Interhemispheric Thermohaline Circulation in a Coupled Box Model

JEFFERY R. SCOTT, JOCHEM MAROTZKE, AND PETER H. STONE

Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

The interhemispheric thermohaline circulation is examined using Rooth's three-box ocean model, whereby overturning strength is parameterized from density differences between high-latitude boxes. Recent results with general circulation models indicate that this is a better analog of the Atlantic thermohaline circulation than a single-hemisphere box model. The results are compared with those of hemispheric box model studies, where possible, and the role of asymmetrical freshwater forcing is explored.

Using both analytical and numerical methods, the linear and nonlinear stability of the model is examined. Although freshwater forcing in the Southern Hemisphere alone governs overturning strength, increasing freshwater forcing in the Northern Hemisphere leads to a heretofore unrecognized instability in the northern sinking branch due to an increasingly positive ocean salinity feedback. If the northern forcing is instead made weaker than the southern forcing, this feedback becomes negative. In contrast, the ocean salinity feedback is always positive in single-hemisphere models. Nonlinear stability, as measured by the size of the perturbation necessary to induce a permanent transition to the southern sinking equilibrium, is also observed to depend similarly on the north–south forcing ratio.

The model is augmented with explicit atmospheric eddy transport parameterizations, allowing examination of the eddy moisture transport (EMT) and eddy heat transport (EHT) feedbacks. As in the hemispheric model, the EMT feedback is always destabilizing, whereas the EHT may stabilize or destabilize. However, in this model whether the EHT stabilizes or destabilizes depends largely on the sign of the ocean salinity feedback and the size of the perturbation. Since oceanic heat transport in the Southern Hemisphere is weak, the Northern Hemisphere EMT and EHT feedbacks dominate.

1. Introduction

The present-day thermohaline circulation (THC) is characterized by deep-water formation in the North Atlantic Ocean and upwelling on a broad scale over much of the world's oceans (Gordon 1986; Macdonald and Wunsch 1996). An hierarchy of models ranging from simple box models to coupled ocean–atmosphere general circulation models suggests that multiple equilibria of the THC are not only possible, but under some conditions alternate circulation patterns are favored (Stommel 1961; Manabe and Stouffer 1988; Marotzke 1990; Marotzke and Willebrand 1991; Hughes and Weaver 1994; Mikolajewicz and Maier-Reimer 1994; Schiller et al. 1997). Stommel's (1961) box model offers the fundamental explanation of multiple steady states in a single hemisphere (e.g., the North Atlantic), but its applicability in addressing the behavior of the global circulation or even that of the whole Atlantic is obviously limited. No comparable paradigm has emerged for the interhemispheric THC; hence, a *conceptual* understanding is essentially lacking, although there have been numerous studies with both global and single-basin numerical models (see references above).

The density structure of the oceans is clearly dominated by temperature, as indicated by pervasive cold deep water, but salinity variations among cold water locations play a role in determining deep-water formation sites—the most apparent example being the lack of deep-water formation in the relatively fresh North Pacific (Warren 1993). Less apparent is the role of freshening in the high-latitude South Atlantic. Only a few studies have considered the notion that the pole-to-pole density difference in the entire Atlantic might be more important for THC strength [as first proposed by Rooth (1982)], in contrast with Stommel's model, whereby overturning is a function of the equator-to-pole density gradient. Using an ocean general circulation model (OGCM), Hughes and Weaver (1994) diagnosed a linear relation between overturning strength and the difference in zonally and vertically averaged steric height (a measure of depth-integrated pressure), measured between the tip of Africa and the northern latitude with maximum surface density, where sinking is concentrated. This relationship also held true for points along the western

Corresponding author address: Jeffery R. Scott, Dept. of Earth, Atmospheric, and Planetary Sciences, MIT, Room 54-1711, Cambridge, MA 02139.
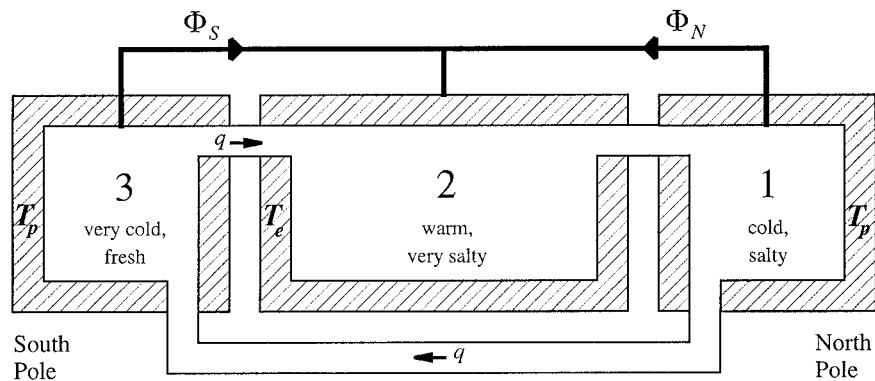E-mail: jscott@mit.edu

FIG. 1. Rooth's conceptual three-box model of thermohaline circulation, showing equilibrium conditions for Northern Hemisphere sinking. The separation between high- and low-latitude boxes is assumed to occur near the peak in atmospheric transports due to baroclinic eddy fluxes, i.e., about 35° latitude.

boundary (A. Weaver 1997, personal communication). In a separate OCGM study, Rahmstorf (1996) found a linear relation between North Atlantic Deep Water (NADW) production and interhemispheric density differences at middepth. An even more robust justification that high-latitude density asymmetry is fundamental in producing an interhemispheric overturning circulation is presented in Klinger and Marotzke (1999); they found that strong cross-equatorial flow was generated in a temperature-only OGCM when high-latitude restoring temperatures were weakly asymmetric.

To capture what we believe is the fundamental underlying dynamics of the interhemispheric circulation, this study adopts Rooth's (1982) box model of a single ocean basin. Unlike Stommel's model, overturning strength in Rooth's model is proportional to density differences between high-latitude boxes. Recent results from Wang et al.'s (1999a) and Tziperman's (1997) OGCM studies showed that, although the global deep-water formation rate is relatively insensitive to freshwater forcing, the formation of NADW was a strong function of high-latitude freshwater forcing. In addition, Wang et al. found increased NADW production if the South Atlantic's density is decreased through a larger poleward moisture flux, in accord with the pole-to-pole conceptual framework, leading to the conjecture that Rooth's model is a better analog than Stommel's model for the NADW branch within the global framework. Although present in the Atlantic, Antarctic Bottom Water is ignored in this picture, as NADW production dominates both water-mass properties and overall mass and heat transports of the Atlantic basin.

Our focus here is primarily on the stability characteristics of the pole-to-pole circulation, explicitly considering the role of north–south asymmetries in freshwater forcing. This issue was addressed in Rahmstorf (1996); a crucial difference, however, is that our study has poleward atmospheric freshwater flux in both hemispheres, as is observed. In contrast, Rahmstorf's flux was equatorward in the Southern Hemisphere, which

was essential to his main findings. We do not explicitly consider cross-basin freshwater transport, which has been proposed by Broecker et al. (1990) as essential in driving the present-day THC but shown as not necessary to reproduce the "conveyor" circulation in OGCMs (Marotzke and Willebrand 1991; Hughes and Weaver 1994).

A second focus is to examine ocean–atmosphere interactions in the context of the pole-to-pole circulation, in particular examining the ocean and atmospheric feedbacks and the behavior of climate transitions. To date, much of our conceptual understanding of how large-scale ocean–atmosphere interactions might influence the THC was derived from simple models representing the climate system in a single hemisphere (Nakamura et al. 1994; Marotzke and Stone 1995, hereafter MS; Lohmann et al. 1996; Marotzke 1996, hereafter M96). Wang et al. (1999b) also investigated the roles played by the atmospheric transports in the different hemispheres; we will discuss these results in a later section.

The paper is organized as follows. In section 2, an extension of Rooth's model is presented with some key analytical results. Section 3 considers the linear stability of the northern sinking equilibrium; this instability does not invoke unrealistic equatorward atmospheric moisture transport and has apparently been overlooked by all previous researchers. Section 4 looks at the critical salt perturbation necessary to induce a permanent transition to the southern sinking circulation. In section 5, atmospheric moisture and heat feedbacks are added to the model and the linear and finite amplitude stability characteristics are readdressed. Section 6 includes a brief summary and considers the implications of these results.

2. Model formulation and equilibrium solutions

Rooth's (1982) three-box model (illustrated schematically in Fig. 1) consists of a single equatorial box (box 2) and two high-latitude boxes (boxes 1 and 3),

interconnected by an insulated deep current pathway of insignificant volume. As in Stommel's model, the boxes are assumed to be well mixed; this assumption precludes the possibility of the "polar halocline catastrophe" (Bryan 1986; Zhang et al. 1993), as convection is not explicitly parameterized. The high-latitude boxes are equal in volume. Freshwater forcing from the Tropics, equivalent to salt fluxes Φ_N and Φ_S out of the high-latitude boxes toward the equatorial box, are taken as constant. Rooth considered the case where the freshwater forcing was identical in each hemisphere; this study, however, explicitly considers the ramifications of asymmetrical freshwater fluxes. The high and low latitude boxes are bathed in thermal reservoirs of temperature T_p and T_e, respectively. All three boxes undergo thermal Newtonian relaxation at the same rate, λ. Rooth did not include explicit atmospheric poleward heat fluxes in his model. However, by choosing reasonable equatorial and polar temperatures rather than values derived solely from the radiation balance, say 30°C and 0°C for T_e and T_p, respectively, one implicitly includes these heat fluxes in the Newtonian relaxation terms (M96).

As in Stommel's (1961) model, the thermohaline overturning strength q is determined from density differences using a linearized equation of state. The distinguishing characteristic of Rooth's model is that q is parameterized as a linear function of the density difference between the two high-latitude boxes:

$$q = k[\alpha(T_3 - T_1) - \beta(S_3 - S_1)], \qquad (1)$$

where α and β are the thermal and haline expansion coefficients, respectively, and k is a hydraulic constant.

With Northern Hemisphere sinking defined as $q > 0$, the governing system of equations for heat and salt conservation is as follows (Rooth 1982):

$$\dot{T}_1 = \begin{cases} q(T_2 - T_1) + \lambda(T_p - T_1), & q \geq 0 \\ |q|(T_3 - T_1) + \lambda(T_p - T_1), & q < 0 \end{cases}$$

$$\dot{T}_2 = \begin{cases} \dfrac{q}{V}(T_3 - T_2) + \lambda(T_e - T_2), & q \geq 0 \\[2mm] \dfrac{|q|}{V}(T_1 - T_2) + \lambda(T_e - T_2), & q < 0 \end{cases}$$

$$\dot{T}_3 = \begin{cases} q(T_1 - T_3) + \lambda(T_p - T_3), & q \geq 0 \\ |q|(T_2 - T_3) + \lambda(T_p - T_3), & q < 0 \end{cases}$$

$$\dot{S}_1 = \begin{cases} q(S_2 - S_1) - \Phi_N, & q \geq 0 \\ |q|(S_3 - S_1) - \Phi_N, & q < 0 \end{cases}$$

$$\dot{S}_2 = \begin{cases} \dfrac{q}{V}(S_3 - S_2) + \dfrac{\Phi_N + \Phi_S}{V}, & q \geq 0 \\[2mm] \dfrac{|q|}{V}(S_1 - S_2) + \dfrac{\Phi_N + \Phi_S}{V}, & q < 0 \end{cases}$$

$$\dot{S}_3 = \begin{cases} q(S_1 - S_3) - \Phi_S, & q \geq 0 \\ |q|(S_2 - S_3) - \Phi_S, & q < 0, \end{cases} \qquad (2)$$

where V is the ratio of the equatorial to the high-latitude box volume. For the northern sinking equilibrium, the high-latitude Northern Hemisphere box will be more salty and warmer than the southern high-latitude box due to the flow of equatorial water into box 1, reflecting that the surface North Atlantic is indeed warmer and more salty than comparable latitudes in the Southern Hemisphere (Levitus 1982). Rooth's model, by construction, is "thermally driven" in that sinking can occur only at the poles, but the equilibrium THC direction (and strength) is determined by the pole-to-pole salinity difference. Due to warm advection into the sinking box, the equilibrium pole-to-pole temperature contrast partially counteracts the haline term in (1). In this regard, the pole-to-pole parameterization is quite different from that in the single-hemisphere models, where the reverse is true: namely, the (equator-to-pole) temperature contrast drives flow and the salinity difference functions as a brake.

To examine solutions to Rooth's model, it is convenient to recast (1) and (2) in terms of the temperature and salinity differences between the equatorial box and the high-latitude boxes: $T_N \equiv T_2 - T_1$, $T_S \equiv T_2 - T_3$, $S_N \equiv S_2 - S_1$, and $S_S \equiv S_2 - S_3$. Physical solutions must have positive T_N, T_S, S_N, and S_S at equilibrium. Setting $\dot{S}_1$ and $\dot{S}_2$ to zero yields, respectively, for $\overline{q} \geq 0$:

$$\overline{S}_N = \frac{\Phi_N}{\overline{q}}, \qquad \overline{S}_S = \frac{\Phi_N + \Phi_S}{\overline{q}} \qquad (3)$$

so that

$$\overline{q} = k\left[\alpha(\overline{T}_N - \overline{T}_S) + \beta\frac{\Phi_S}{\overline{q}}\right] \qquad (4)$$

(the overbar denotes an equilibrium value). From (4), it is possible to solve for $\overline{q}$:

$$\overline{q} = \frac{\alpha k(\overline{T}_N - \overline{T}_S)}{2} \pm \frac{1}{2}\sqrt{k^2\alpha^2(\overline{T}_N - \overline{T}_S)^2 + 4k\beta\Phi_S} \qquad (5)$$

(Rahmstorf 1996). Because the temperature tendency equations do not contain S_N or S_S, the equilibrium THC depends on the freshwater forcing *in the upwelling hemisphere only* (hereafter, we will refer to the sinking and upwelling hemispheres as the Northern and Southern Hemisphere, respectively, with the understanding that the roles are reversed for the southern sinking equilibrium). Physically, the salinity "transformation" between the inflow and outflow of box 3 depends on Φ_S and salinity advection by the mean overturning; the overturning strength, in turn, is a function of the salinity transformation. Hence, Φ_N is not needed to close the system, and subsequently falls out of the equilibrium solution.

Manipulating the temperature tendency equations to find $\overline{T}_N$ and $\overline{T}_S$ is less amenable to simple analytical methods; therefore the equilibrium solution is found numerically. For the northern sinking solution, $\overline{T}_N < \overline{T}_S$ so that only the positive root in (5) results in a physically meaningful value. Thus, there is only one equilibrium

TABLE 1. Model parameters.

Parameter	Symbol	Value
Thermal expansion coefficient	α	1.5×10^{-4} K^{-1}
Haline expansion coefficient	β	0.8×10^{-3} psu^{-1}
High-latitude thermal bath	T_p	0°C
Equatorial thermal bath	T_e	30°C
Newtonian relaxation constant	λ	12.9×10^{-10} s^{-1}
Hydraulic constant	k	1.5×10^{-6} s^{-1}
Equatorial/high-latitude box volume ratio	V	2
Ocean basin size		60° wide, 5000 m deep

TABLE 2. Equilibrium results for the three cases of freshwater forcing mentioned in the text.

	Case I ($\Phi_N = \Phi_S$)	Case II ($\Phi_N = \frac{1}{2}\Phi_S$)	Case III ($\Phi_N = \frac{3}{2}\Phi_S$)
$\bar{S}_1$ (psu)	34.85	35.00	34.69
$\bar{S}_2$	35.46	35.31	35.61
$\bar{S}_3$	34.23	34.39	34.08
$\bar{S}_N$	0.61	0.31	0.92
$\bar{S}_S$	1.22	0.92	1.53

solution for $\bar{q} > 0$, for which the stability has not yet been determined. The single-hemisphere model allows the possibility of two high-latitude sinking solutions (Stommel 1961): a solution with a strong THC but a weak salinity gradient (stable) and a solution with a weak THC and a large salinity gradient (unstable). Our simpler result is due to a single permissible value for $\bar{S}_N$ and $\bar{S}_S$ at equilibrium. Equation (5) also suggests that for large Φ_S, the equilibrium thermohaline overturning strength is nearly proportional to $\sqrt{\Phi_S}$. This result was first suggested by Rooth, who considered the circulation that developed if $\bar{T}_N = \bar{T}_S$ (i.e., given strong Newtonian cooling with symmetric polar temperature baths), although Rooth did not recognize that flow strength depended only on the Southern Hemisphere forcing.

Given the possibility of asymmetric freshwater forcing, it is worthwhile to examine several limiting cases for Φ_N and Φ_S. First, consider the resulting steady state when Φ_N is decreased while holding Φ_S constant. The strength of the THC, and therefore the temperature structure, remains unchanged, but the salinity gradients $\bar{S}_N$ and $\bar{S}_S$ become smaller (in fact, $\bar{S}_N \rightarrow 0$ as $\Phi_N \rightarrow 0$). In other words, changes in freshwater forcing in the Northern Hemisphere result in a redistribution of salt such that the overall $\bar{q}$ is unaffected. Conversely, if Φ_N is increased, the salinity gradients are larger, again without affecting $\bar{q}$, $\bar{T}_N$, or $\bar{T}_S$. Next, consider decreased freshwater forcing in the Southern Hemisphere. In this case, the steady-state THC is weaker, and $\bar{S}_N$ and $\bar{S}_S$ both increase. This weakening in $\bar{q}$ results in equilibrium temperatures closer to the respective thermal reservoirs. For $\Phi_S \ll \Phi_N$, the THC must be very weak and therefore $\bar{S}_N$ and $\bar{S}_S$ large, since the freshwater forcing in the Southern Hemisphere is the sole process by which box 3 can be less dense than box 1. Conversely, if Φ_S is increased, a greater overturning strength and smaller salinity gradients result.

Rooth also discussed the possibility of a no-flow, transient solution for symmetric freshwater forcing where $T_1 = T_3$ and $S_1 = S_3 = -S_2 = -\Phi t$. This solution is not physically realistic as $t \rightarrow \infty$, as it is unstable to small perturbations. This type of symmetric circulation has also been found to be unstable in models of various complexity (e.g., Bryan 1986; Marotzke 1990). However, the solution is important in the transient response

of the system to large perturbations, and will be discussed in more detail in subsequent sections.

To analyze the stability behavior numerically, it is necessary to choose an appropriate set of parameters (see Table 1). These choices are similar to those in M96 and are thought to reflect a hypothetical ocean basin on Earth. The value for the hydraulic constant is selected so that the cross-equatorial flow will roughly agree with the 14 Sv (Sv $\equiv 10^6$ m^3 s^{-1}) estimate of Macdonald and Wunsch (1996). The choice of λ is a function of the efficiency of longwave radiation, the efficiency of atmospheric heat transport, and the ocean basin size. An equation for λ, as well as a value for the implicit atmospheric heat flux efficiency, is given in M96. Note, however, that this equation applies to the temperature difference between the two boxes. When applied to individual boxes in Rooth's model, this choice of λ ensures that temperature differences between boxes will be damped back to equilibrium at a realistic rate, but the global mean temperature will be damped too rapidly, about five times faster than the "global" restoring timescale calculated in MS. This discrepancy would be problematic for global warming experiments, but does not affect results presented here.

Three choices for the freshwater forcing are considered: a symmetric case (I), where $\Phi_N = \Phi_S = 0.9 \times 10^{-10}$ psu s^{-1}, corresponding to a flux of 0.27 Sv (or equivalently, 0.41 m yr^{-1}) of net precipitation/runoff into each high-latitude box based on a reference salinity of 35 psu; an asymmetric case (II), where Φ_N is decreased 50% and Φ_S held fixed; and an asymmetric case (III), where Φ_N is increased 50%. Case III is most like the present Atlantic basin because most of North America and a large part of Asia drain into the North Atlantic, whereas the South Atlantic receives little river runoff at high latitudes (Baumgartner and Reichel 1975; Broecker et al. 1990). Broecker's estimate (based on Baumgartner and Reichel's climatology data) of total meridional moisture transport across 40°, including precipitation minus evaporation, is 0.31 Sv to the North Atlantic but only 0.18 Sv to the South Atlantic. By holding Φ_S fixed in these experiments, an identical overturning strength of $\bar{q} = 1.47 \times 10^{-10}$ s^{-1}, or about 16 Sv, is produced in all cases. The equilibrium temperatures are $T_1 = 2.9$°C, $T_2 = 28.4$°C, and $T_3 = 0.3$°C. The equilibrium salinities are presented in Table 2. The

pole-to-pole salinity difference is identical in all three cases, as required for identical overturning, although the equator-to-pole salinity differences are greatest in case III. The equilibrium solutions for all three cases are stable to small perturbations.

3. Linear stability analysis

a. Perturbation tendency equations

In order to address the stability of the three equilibria listed in Table 2, consider a perturbation analysis of the governing equations, keeping only the highest order terms:

$$\dot{T}'_N = -q'\left(\overline{T}_N + \frac{\overline{T}_S}{V}\right) - \overline{q}\left(T'_N + \frac{T'_S}{V}\right) - \lambda T'_N \tag{6a}$$

$$\dot{T}'_S = -q'[-\overline{T}_N + \overline{T}_S(1 + 1/V)]$$
$$\quad - \overline{q}[-T'_N + T'_S(1 + 1/V)] - \lambda T'_S \tag{6b}$$

$$\dot{S}'_N = -q'\left(\overline{S}_N + \frac{\overline{S}_S}{V}\right) - \overline{q}\left(S'_N + \frac{S'_S}{V}\right) \tag{6c}$$

$$\dot{S}'_S = -q'[-\overline{S}_N + \overline{S}_S(1 + 1/V)]$$
$$\quad - \overline{q}[-S'_N + S'_S(1 + 1/V)] \tag{6d}$$

(a prime denotes a deviation from equilibrium). The perturbation temperature tendency equations (6a) and (6b) include three terms, from left to right: advection of the mean temperature by anomalous flow, advection of temperature anomalies by the mean flow, and a Newtonian cooling term. Note that (6a) is nearly identical to the comparable equation from the single-hemisphere model [cf. Eq. (40) in M96], except that now the influence of the opposite hemisphere is present in the mean and anomalous flow terms. The salinity tendency equations (6c) and (6d) are similar to (6a) and (6b), except that salinity rather than temperature is advected by mean and anomalous flow, and there is no analogous restoring term.

Four processes can be identified in (6a–d) that affect stability (as discussed in M96): the mean flow, Newtonian cooling, the ocean heat transport feedback, and the salinity transport feedback. In isolation, the mean flow would eliminate temperature and salinity anomalies on a flushing timescale, the time for the mean flow to recycle the water in a given box. This timescale is 215 yr for a high-latitude box and 430 yr for the equatorial box. Newtonian cooling acts in conjunction with the mean flow to restore temperature even more rapidly than salinity, particularly since $\lambda \gg \overline{q}$.

In the single-hemisphere model, determining the sign of the ocean feedbacks is straightforward. An increase in temperature (or salinity) in the high-latitude box due to anomalous flow corresponds with a similar decrease in the equatorial box; for example, a positive flow anomaly advects more warm, salty water to the high-latitude

box and more cold, fresh water into the equatorial box. Since warm advection weakens and salt advection strengthens flow, the ocean temperature feedback is negative and the salinity feedback is positive.

In Rooth's model, the feedback analysis is more complicated. The outflow from the sinking high-latitude box travels immediately to the opposite high-latitude box, whereas the inflow is from the equatorial box. In order to determine the sign and/or importance of the feedbacks in Rooth's model, one must consider how perturbations in ocean circulation affect the properties of both high-latitude boxes. For the temperature feedback, this assessment remains straightforward: A negative flow anomaly leads to strong cooling in the Northern Hemisphere (tending to strengthen flow) and weak cooling in the Southern Hemisphere (tending to weaken flow further). Because of the large disparity in heat flux between the two hemispheres, only the stabilizing Northern Hemisphere feedback is important, so the ocean temperature feedback essentially functions as it does in Stommel's model.

The ocean salinity feedback is the more interesting of the two, which we will see is fundamental in understanding the behavior of Rooth's model. For symmetric forcing, both high-latitude boxes in equilibrium must have identical advective salt transport [see Eq. (2)]. Thus, if a negative flow anomaly is introduced there is (initially) no salinity feedback. For weaker freshwater forcing in the north (e.g., case II), $\overline{S}_N < \overline{S}_S - \overline{S}_N$, and a negative flow anomaly leads to greater freshening in the southern high-latitude box and hence increased flow, an overall negative feedback. Conversely, for stronger freshwater forcing in the north (e.g., case III), $\overline{S}_N > \overline{S}_S - \overline{S}_N$, and a negative flow anomaly leads to greater freshening in the northern high-latitude box and further decreased flow, a positive feedback. Hence, the sign of the model's ocean salinity feedback depends on asymmetries in freshwater forcing.

b. Analytical solution for prescribed temperatures

To clarify the relationship between freshwater forcing and stability, we compute the eigenvalues analytically for the system (6c) and (6d) assuming constant temperature; that is, $T'_N = T'_S = 0$, equivalent to taking the limit $\lambda \to \infty$. For simplicity, first consider the case where $\overline{T}_N - \overline{T}_S \to 0$. By substituting $q' = -k\beta(S'_N - S'_S)$ into (6c) and (6d), the system becomes

$$\dot{S}'_N = k\beta\{[2\overline{S}_N + (-1 + 1/V)\overline{S}_S]S'_N$$
$$\quad + [(-1 + 1/V)\overline{S}_N + (-2/V)\overline{S}_S]S'_S\} \tag{8a}$$

$$\dot{S}'_S = k\beta\{[-2\overline{S}_N + (2 + 1/V)\overline{S}_S]S'_N$$
$$\quad + [(2 + 1/V)\overline{S}_N + (-2 - 2/V)\overline{S}_S]S'_S\}. \tag{8b}$$

This system can be rewritten in the form

$$\begin{pmatrix} \dot{S}'_N \\ \dot{S}'_S \end{pmatrix} = k\beta\mathbf{A}\begin{pmatrix} S'_N \\ S'_S \end{pmatrix}, \tag{9}$$

where the elements of the matrix $\mathbf{A}$ are given by the terms in brackets in (8). It is straightforward to show that unless $\Phi_S \ll \Phi_N$, the eigenvalues of $\mathbf{A}$ are complex. Hence, the real part of the complex conjugate pair is simply one-half the trace of $\mathbf{A}$, which is given by

$$\text{trace}(\mathbf{A}) = \left(4 + \frac{1}{V}\right)\overline{S}_N + \left(-3 - \frac{1}{V}\right)\overline{S}_S. \tag{10}$$

Instability occurs when $\text{trace}(\mathbf{A}) > 0$. Substituting $\overline{S}_N$

and $\overline{S}_S$ from (3) into (10), we are left with a rather simple statement regarding the model's stability:

$$\frac{\Phi_S}{\Phi_N} < \frac{1}{(3 + 1/V)} \tag{11}$$

produces a growing mode. Unless V is small, the equilibrium becomes unstable when the Southern Hemisphere freshwater forcing is approximately less than one-third of the Northern Hemisphere value.

In the more general case where $\lambda \to \infty$ but $\overline{T}_N - \overline{T}_S$ does not vanish, one can show this condition becomes

$$\frac{\Phi_S}{\Phi_N} < \frac{k\beta\Phi_S}{k\alpha(\overline{T}_N - \overline{T}_S)(1 + 1/2V)[k\alpha(\overline{T}_N - \overline{T}_S) + \sqrt{k^2\alpha^2(\overline{T}_N - \overline{T}_S)^2 + 4k\beta\Phi_S}] + k\beta\Phi_S(3 + 1/V)}. \tag{12}$$

The neutrally stable curve for our equilibria $\overline{T}_N - \overline{T}_S$ is computed using (12) and is shown in Fig. 2. Along a horizontal line (i.e., with a constant south:north forcing ratio), the equilibrium becomes unstable as it crosses the neutrally stable curve from right to left. Also plotted is the isopleth $\Phi_S = 0.9 \times 10^{-10}$ psu s^{-1}, representing the family of equilibria with constant flow strength (and, consequently, mean flow feedback strength) as given in section 2. As the Φ_S isopleth crosses this neutrally stable curve from above, the equilibrium changes from stable to unstable. (As it happens, the intersection point is nearly identical to the case III configuration.) In other words, given the equilibrium temperature structure it is also possible for the solution to become unstable by

weakening the freshwater forcing in both hemispheres, as long as $\Phi_S/\Phi_N < 1$, without changing the asymmetry forcing ratio.

A comparison between neutrally stable curves in Fig. 2 suggests that the equilibrium temperature structure leads to a reduced area of stability in parameter space. Recall the equilibrium temperature structure acts as a brake, producing a less robust overturning and hence weakening the mean flow feedback. Conversely, rapid restoring to a temperature structure that helps drive flow would be expected to lead to a greater domain of stability.

c. Perturbation experiment

In section 3b we examine a simplified version of the model for analytical convenience. Here, we will compare the behavior of the simplified model and the full model, which admits the stablizing ocean temperature feedback, using numerical methods.

In order to examine how the model's feedbacks affect the response to a flow anomaly, we perturb the equilibrium state by adding a -0.001 salt anomaly to the northern high-latitude box, sufficiently small so as to avoid nonlinear effects. The time evolution is shown in Fig. 3. In case I, all feedbacks are nonpositive, and the system rebounds rapidly following the perturbation. The positive salinity feedback in case III only marginally interferes with the system's ability to reequilibrate. However, if the stabilizing ocean temperature feedback is removed by fixing temperatures (equivalent to letting $\lambda \to \infty$ for relaxation back to the equilibrium box temperatures rather than the thermal reservoir temperatures, as considered in section 3b), the asymmetry in freshwater forcing leads to a rather dramatic response: the system oscillates about the equilibrium flow strength, very weakly damped (see Table 3), for this choice of
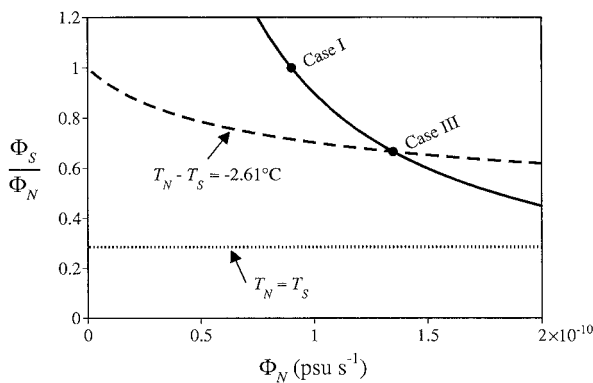
FIG. 2. Stability diagram expressed in parameter space, in the simplifying limit as $\lambda \to \infty$. The abscissa represents increases in the freshwater forcing in the Northern Hemisphere (in psu s^{-1}), and the ordinate expresses the ratio of the Southern Hemisphere to Northern Hemisphere freshwater forcing. Solid curve: isopleth given by $\Phi_S = 0.9 \times 10^{-10}$ psu s^{-1}; dashed: neutrally stable curve with temperatures held constant at their equilibrium values; dotted: neutrally stable curve with temperatures held constant at $T_N = T_S$. For $\Phi_S = 0.9 \times 10^{-10}$ psu s^{-1} (solid curve), all equilibria have the same flow strength irrespective of Φ_N, but they are unstable below the dashed curve.
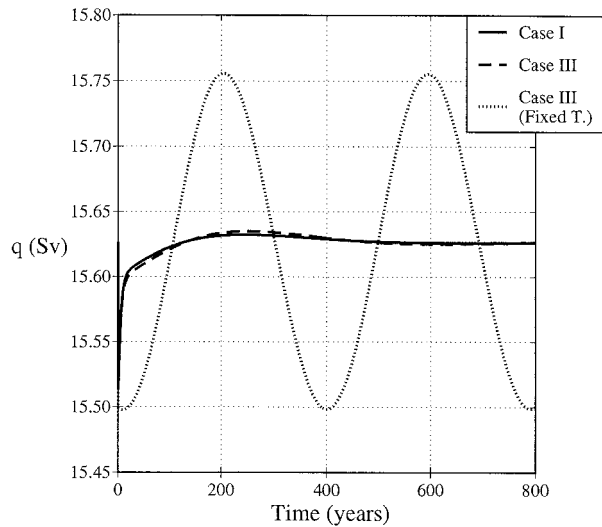
FIG. 3. Time evolution of overturning strength following a salinity perturbation of −0.001 psu in box 1. Solid: case I, symmetric freshwater forcing; dashed: case III, stronger Northern Hemisphere freshwater forcing; dotted: case III freshwater forcing with temperatures held at their equilibrium values.

freshwater forcing. This behavior is consistent with the analytical results plotted in Fig. 2, where it is noted that the case III equilibrium is situated very close to the pertinent neutrally stable curve.

The physical mechanism for this oscillation is linked to the system's salinity response to flow anomalies. Following a negative flow perturbation, the equatorial box becomes more salty while the high-latitude boxes are freshened. Except for the mean flow feedback, the system has no direct means by which it can inhibit this anomalous equatorial salt buildup because overturning strength is not parameterized as a function of the equatorial properties. As the anomalous salt is advected away from the Tropics by the mean flow to the north, the salinity in box 1 is restored more rapidly than box 3. Consequently, q overshoots its equilibrium value (a modest overshooting is also apparent in Fig. 3 for both strongly damped responses). An alternative way to view this mechanism is to consider that an anomaly in flow produces a salt imbalance between boxes 2 and 3. To restore equilibrium, the imbalance must be advected

through box 1, as there is no southward flow path from box 2 to box 3 [this is similar to the delay mechanism discussed in Welander's (1986) "oscillator gallery," examining possible explanations of THC variability]. This overshooting behavior was also observed in the Saravan and McWilliams (1995) two-dimensional ocean model, although the authors did not explicitly address the cause of this behavior.

A calculation of the eigenvalues of the linearized equations is presented in Table 3. If the temperature feedback is included, there are two real and two complex eigenvalues. These real eigenvalues reflect damping timescales on the order of λ^{-1}. Damping is only slightly weaker in case III than case I. However, when temperature is held constant, only the complex eigenvalues remain, which are linked to the salinity equations and the aforementioned oscillatory mechanism (thus the oscillation has a period corresponding to the advective timescale). Here, the damping in case III is three orders of magnitude smaller than in case I even though the period of oscillations, about 700 yr, changes little. In case I this damping timescale is approximately the flushing timescale, which is what one would expect given the neutral salinity feedback.

4. Critical perturbation experiments

Given a salt perturbation of significant magnitude, the northern sinking circulation will reverse to a southern sinking circulation if this reverse circulation is stable. The size of the critical perturbation required to induce this transition is a second test of an equilibrium's stability (M96), addressing its nonlinear behavior. This stability test is commonly employed in OGCM studies, simulating a large meltwater discharge (e.g., Bryan 1986; Manabe and Stouffer 1995).

Although the equilibrium temperature structure initially facilitates a switch to a southern sinking—it is not necessary to make box 1 as fresh as box 3 because box 1 is warmer, hence permitting a reverse flow driven by the temperature difference—the thermal structure provides resistance against a permanent transition. The onset of southern sinking advects warm water to the southern box, whereas the northern box is cooled, helping to restore the northern sinking equilibrium. This behavior

TABLE 3. Eigenvalues and associated damping timescales for the linearized perturbation equations. Results are for symmetric freshwater forcing (case I) and stronger Northern Hemisphere freshwater forcing (case III). For the upper set, the box temperatures are allowed to vary normally. In the last row, the box temperatures were held constant at their equilibrium values.

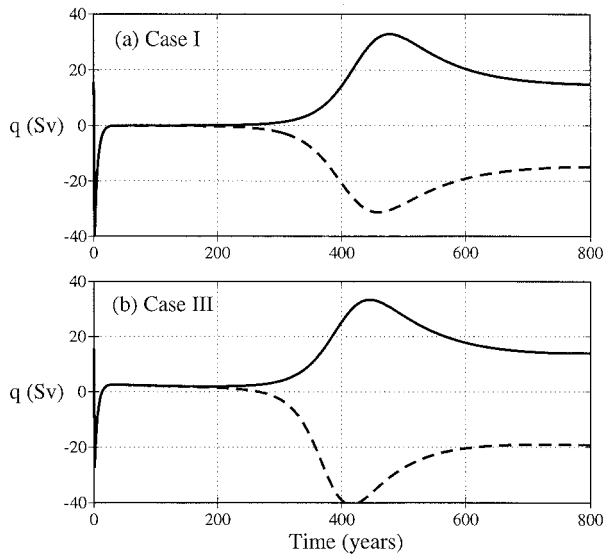|	Case I		Case III	
	Eigenvalue	Damping time scale (years)	Eigenvalue	Damping time scale (years)
With temperature feedback:	-6.8×10^{-9}	5	-6.5×10^{-9}	5
	-1.3×10^{-9}	25	-1.3×10^{-10}	25
	$-2.0 \times 10^{-10} \pm 2.4 \times 10^{-10}i$	162	$-1.6 \times 10^{-10} \pm 2.8 \times 10^{-10}i$	204
Temperature fixed:	$-1.8 \times 10^{-10} \pm 4.8 \times 10^{-10}i$	172	$-3.3 \times 10^{-13} \pm 5.1 \times 10^{-10}i$	95 734

FIG. 4. Time evolution of overturning strength following a salinity perturbation in box 1. Solid: response following a subcritical perturbation; dashed: response following a critical perturbation. (a) Case I, symmetric freshwater forcing, with a critical perturbation of -0.522 psu; (b) case III, stronger Northern Hemisphere freshwater forcing, with a critical perturbation of -0.387 psu.

is illustrated in Fig. 4a, which shows the time evolution of the case I equilibrium following a sub- and supercritical perturbation. Initially, strong southern sinking is induced, but it is weakened by the subsequent advection of heat (this observed rebound is entirely absent if temperatures are held constant, and the critical perturbation is subsequently much smaller). This brief reverse circulation also advects salinity from the equatorial box to the south; to induce a permanent transition, enough salt needs to be advected into the southern box during this period.

Following the initial rebound, the system persists in a state of extremely weak flow for about 200 years prior to the bifurcation of trajectories. During this period, the properties of the high-latitude boxes remain nearly identical, experiencing similar freshening and thermal damping. This state is Rooth's unstable no-flow transient solution, which here acts like a saddle point. The critical perturbation behavior can be summed up as follows. If the temperature rebound pushes the system just across the no-flow solution (which acts like an unstable manifold), the original northern sinking equilibrium will eventually be recovered. If the rebound is too weak, the final outcome is the southern sinking equilibrium.

An analysis of the linear stability about this intermediate solution is helpful in understanding the transient behavior. Initially, Newtonian cooling and an ocean temperature feedback are both strongly stabilizing, as they tend to weaken any flow anomaly. Thus, the system is attracted to this solution, given any finite amplitude perturbation that sufficiently reduces the pole-to-pole salinity difference. Once attracted to the solution, how-

ever, a positive salinity feedback grows as the equator-to-pole salinity difference increases, ultimately causing the system to branch away.

In the absence of Newtonian cooling, an estimate of the timescale for this bifurcation to occur can be obtained by comparing the magnitude of $-\alpha\overline{T}$ to $\beta\overline{S}$, where $\overline{T}$ and $\overline{S}$ are the equator-to-pole temperature and salinity differences, respectively. When summed, these terms can be interpreted as the high-latitude density tendency induced by anomalous flow from the equator. When $\beta\overline{S}$ is large, which occurs when $\overline{S} > 6$ psu, the positive salinity feedback is dominant, and a positive salinity anomaly in the Northern Hemisphere drives an increasing overturning. The observed bifurcation occurs before $\overline{S}$ becomes quite this large, however, as Newtonian cooling actually weakens the ocean temperature feedback (in a manner analogous to that by which the atmospheric heat transport feedback weakens the ocean temperature feedback, as described in section 5). It is interesting to note that Newtonian cooling, which initially helps attract the thermally driven system to the no-flow solution, ultimately assists in driving it away.

For asymmetrical freshwater forcing, the no-flow solution is not possible per se due to the different freshening rates of the high-latitude boxes. However, an analogous "quasi-no-flow" solution does occur, as shown for case III in Fig. 4b. Here, a weak flow is necessary to offset the greater freshening in the north, so that the high-latitude density contrast remains nearly constant (flow actually weakens slightly over time—the buildup of salt in box 2 implies that weaker advection is necessary to maintain the quasi-balance). For greater freshening in the north, a smaller critical perturbation is necessary since the quasi-no-flow solution is "closer" in phase space. Conversely, for increased freshening in the south, the critical perturbation is larger. Hence, both the linear and finite amplitude stability criteria exhibit the same qualititative results for asymmetrical freshwater forcing.

A summary schematic of the model's behavior for increased freshening in the north is shown in Fig. 5. As discussed in section 2, flow strength of the northern sinking branch is independent of Φ_N, whereas the southern sinking branch goes approximately as $\sqrt{\Phi_N}$. As Φ_N increases, the quasi-no-flow overturning strength must increase to offset the asymmetry in forcing, eventually converging with the stable equilibrium branch as a subcritical Hopf bifurcation; that is, the equilibrium is neutrally stable at this point. As a result, the critical perturbation goes to zero at the bifurcation. This behavior is akin to that exhibited in the single-hemisphere model, where the unstable and stable high-latitude sinking equilibria merge at a critical value of freshwater forcing. Unlike the single-hemisphere model, though, the northern sinking equilibrium exists, albeit unstable, for freshwater forcing above the limit of neutral stability.

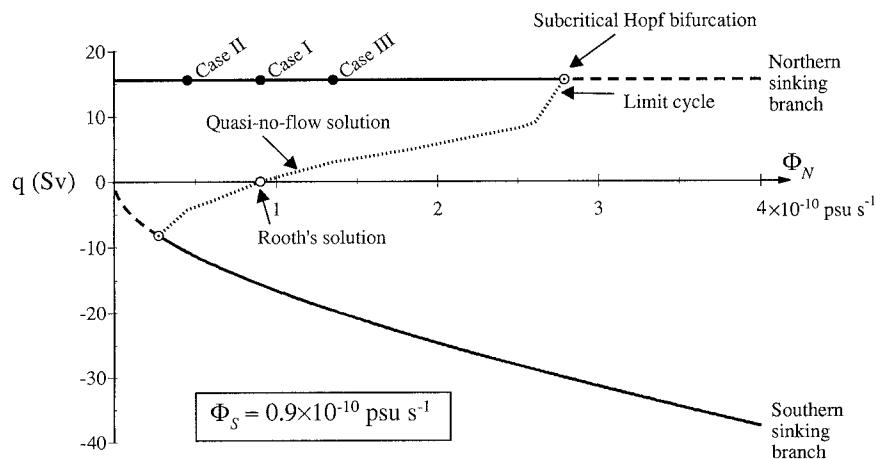The phase space depicted in Fig. 5 depends on the value of Φ_S, which governs the mean flow strength of

FIG. 5. Bifurcation diagram of Rooth's model given constant freshwater forcing in the Southern Hemisphere, showing flow strength as a function of the Northern Hemisphere forcing. Solid: stable portions of equilibrium branches; dashed: unstable portions of equilibrium branches; dotted: transient, unstable intermediate solution observed in critical perturbation experiments (this curve is not rigorously defined, as the flow strength of this intermediate state is not constant).

the northern sinking branch, and, by symmetry arguments, the stability properties of the southern sinking branch; as shown, this figure represents a two-dimensional slice through a three-dimensional phase space. If Φ_S were increased, for instance, the position of the northern sinking line would shift to a larger value of q, although the position of southern sinking solution branch would remain unchanged. The position of the intermediate solution would shift to the right, including

the location of both Hopf bifurcations, that is, increasing the domain of stability for the northern sinking branch but reducing the domain of stability for the southern sinking branch.

The behavior of the unstable limit cycle associated with the Hopf bifurcation is illustrated in Fig. 6, which shows the time evolution of a sub- and supercritical perturbation. The subcritical case evolves as a slowly damped oscillation (the physical mechanism for this oscillation is discussed in section 3), whereas the supercritical case grows slowly prior to a permanent transition to southern sinking. Note that, along most of the quasi-no-flow solution branch, oscillations are not observed, and the behavior is as previously discussed in this section. Near the bifurcation, however, the time span during which this behavior is exhibited contracts (this trend is already apparent in a comparison of Figs. 4a and 4b), and an increasing number of oscillations occurs prior to the transition as the bifurcation is approached. At the bifurcation, all that remains of the no-flow solution behavior is the nadir of the first oscillation.

Through symmetry arguments, a similar Hopf bifurcation occurs in the southern sinking branch, as shown in Fig. 5.

5. Atmospheric moisture and heat transport feedbacks

a. Model reformulation

In this section, explicit atmospheric moisture and heat transports are added to Rooth's model using simple parameterizations (the reader is referred to M96 for a more rigorous description). These parameterizations are based on the observations that the meridional transport into the high-latitude boxes is due primarily to baroclinic
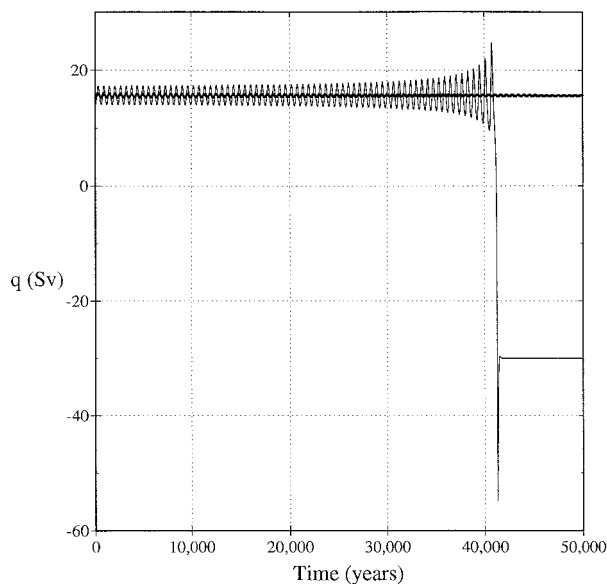

FIG. 6. Response in overturning strength following a supercritical salt perturbation (light solid line) of -0.03 psu and a subcritical perturbation (heavy solid line) of -0.01 psu. The model's Northern Hemisphere freshwater forcing parameter choice is sufficiently large $(2.781 \times 10^{-10}$ psu s$^{-1})$ so that the northern sinking equilibrium is only marginally stable.

FIG. 7. Schematic depicting the model's physical representation of the high-latitude freshwater forcing. Moisture transport occurs from the low-latitude ocean to high latitudes (gray arrows), falling as rain. Some of the rain over land forms rivers draining into the high-latitude ocean box, while other rivers flow southward and drain into tropical ocean waters. The fractional area of the globe represented by the ocean is ε; the fractional area ε_c represents the net catchment area that drains into the high-latitude ocean.

eddies, which depend on the equator-to-pole temperature gradients. The justification for these schemes is both theoretical (Stone and Yao 1990) and empirical (Stone and Miller 1980).

To incorporate an explicit dynamical eddy heat transport feedback (hereafter EHT) into the temperature tendency equations in (2), the Newtonian cooling terms are replaced as follows:

$$\text{in } \dot{T}_1 \quad \lambda(T_p - T_1) \rightarrow \lambda_{\text{rad}}(T_{p_{\text{rad}}} - T_1) + \frac{\chi_{n_N}}{\varepsilon}(T_2 - T_1)^n$$
$$\text{(13a)}$$

$$\text{in } \dot{T}_2 \quad \lambda(T_e - T_2) \rightarrow \lambda_{\text{rad}}(T_{e_{\text{rad}}} - T_2) - \frac{\chi_{n_N}}{\varepsilon V}(T_2 - T_1)^n$$

$$\qquad - \frac{\chi_{n_S}}{\varepsilon V}(T_2 - T_3)^n \qquad \text{(13b)}$$

$$\text{in } \dot{T}_3 \quad \lambda(T_p - T_3) \rightarrow \lambda_{\text{rad}}(T_{p_{\text{rad}}} - T_3) + \frac{\chi_{n_S}}{\varepsilon}(T_2 - T_3)^n,$$
$$\text{(13c)}$$

where n is the heat transport power law relation; χ_{n_N} and χ_{n_S} are heat transport efficiency constants for the Northern and Southern Hemispheres, respectively, for a given n; and ε is the ratio of ocean area to total area (see Fig. 7). The linear relaxation terms in (13a–c) are now entirely radiative, with $T_{e_{\text{rad}}}$ and $T_{p_{\text{rad}}}$ constants implied by a purely radiative balance, 53.5°C and −23°C, respectively, and λ_{rad} is a longwave radiation coefficient. The dynamical heat flux term varies inversely with ε; given a zero heat capacity for the land and the atmosphere, all the atmospheric heat flux is funneled into the ocean, so a model with a smaller ocean fraction is more sensitive to changes in atmospheric heat transport. In our model configuration, ε is assumed to be identical

in the Northern and Southern Hemispheres. For simplicity, the parameterized atmospheric heat transport includes both sensible and latent heat transport. A more sophisticated (and correct) approach would be to parameterize the latent heat transport as a function of the moisture transport.

Similarly, the salinity tendency equations in (2) can be modified to incorporate the eddy moisture transport feedback (hereafter EMT) by replacing the virtual salinity fluxes as follows:

$$\Phi_N \rightarrow \frac{\varepsilon_{c_N}}{\varepsilon} \frac{S_0}{D} \gamma_{m_N}(T_2 - T_1)^m, \qquad \text{(14a)}$$

$$\Phi_S \rightarrow \frac{\varepsilon_{c_S}}{\varepsilon} \frac{S_0}{D} \gamma_{m_S}(T_2 - T_3)^m, \qquad \text{(14b)}$$

where S_0 is a reference salinity of 35 psu, D is the ocean depth, and m, γ_{m_N}, and γ_{m_S} are defined similarly to n, χ_{n_N}, and χ_{n_S} except for eddy moisture transport. All moisture transported to high latitudes is assumed to fall as precipitation over the ocean or over land, returning to the ocean via river flow. Using an analogous argument as before, the eddy moisture flux term also varies inversely with the ocean area fraction. In the model we also allow the possibility that some of these rivers may flow southward and drain into tropical waters, so it is necessary to introduce ε_c, a ratio of the catchment area to the total area (Fig. 7). The eddy moisture flux forcing term is then directly proportional to this catchment ratio. We prescribe ε_c separately for the Northern and Southern Hemispheres (hence the second subscript), which permits asymmetries in the configuration of river networks.

The inclusion of these feedbacks does not modify the representation of flow strength given by (5), except for the substitution of (14b). As before, only the Southern Hemisphere freshwater forcing determines the overturning strength. The magnitude of the Southern Hemisphere temperature gradient is now important, whereas before the temperature dependence was restricted to the pole-to-pole temperature difference $\overline{T}_N - \overline{T}_S$. Also note that $\overline{q}$ is a function of the moisture power law m, but the heat transport power law only indirectly affects the overturning strength through the determination of $\overline{T}_N$ and $\overline{T}_S$.

To understand how these atmospheric feedbacks affect stability for identical mean states, it is necessary to chose appropriate transport efficiency coefficients in (13a–c) and (14a,b), so that the transports in each hemisphere are independent of m and n for the mean (reference) state:

$$\chi_n \overline{T}^n = \text{const}, \qquad \gamma_n \overline{T}^n = \text{const}. \qquad \text{(15a,b)}$$

For the case $m = n = 1$, we assume the heat and moisture efficiencies are symmetric about the equator (thus $m = n = 1$ defines our mean state, which is in spirit close to the linear heat flux of Rooth's model, but now

TABLE 4. Coupled model parameters.

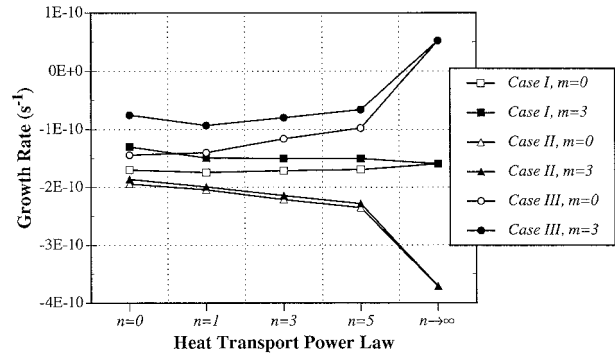| Parameter	Symbol	Value
Radiative damping rate	λ_{rad}	5.1×10^{-10} s^{-1}
Atmospheric heat transport efficiency	χ_1	6.5×10^{-11} s^{-1}
Atmospheric moisture transport efficiency	γ_1	1.663×10^{-11} s^{-1}
NH catchment area : total area	ε_{C_N}	0.5
SH catchment area : total area	ε_{C_S}	0.445

FIG. 8. Largest eigenvalue of the linearized system given varying atmospheric heat and moisture flux power laws. Open symbols: no EMT feedback; solid symbols: cubic atmospheric moisture power law. Abscissa shows increasing temperature damping by the atmospheric heat flux, ranging from fixed heat transport (purely radiative damping of anomalies) to the limiting case of fixed temperature gradients (infinitely strong damping of anomalies).

also permits a varying moisture flux). Varying m and n introduces a slight asymmetry in transport efficiency.

A summary of the new parameters is presented in Table 4. Parameters λ_{rad}, χ_1, and ε_{C_N} are assigned based on similar choices in M96; γ_1 and ε_{c_S} are tuned so that the moisture fluxes exactly match that of case I. Note that $\varepsilon_{c_N} \neq \varepsilon_{c_S}$, as the larger Southern Hemisphere equator-to-pole temperature difference requires a smaller catchment area to produce an identical freshwater forcing. Other properties of this new equilibrium are similar, but not identical, to the comparable equilibrium in section 2 (e.g., overturning strength is reduced from 16 to 14 Sv). Although an exact retuning is possible, it would have required unphysical asymmetries in the radiative processes. To examine stability using the case II and case III freshwater forcings, simply halving or increasing ε_{c_N} by 50%, respectively, is appropriate. Modifications to ε_{c_N} can be interpreted physically as changes to the North Atlantic catchment area.

b. Atmospheric feedbacks

In section 3 a linear stability analysis of the model produced two real, large negative eigenvalues linked to the temperature tendencies and a complex conjugate pair linked to the salinity equations. The real eigenvalues are more negative as n is increased in the reformulated model because the negative feedback on the temperature gradient increases, with only a weak m dependence. Of greater interest is the behavior of the real component of the complex pair, indicating the trend in overall stability. To address this behavior, first consider the response to a small negative anomaly in flow:

$$q' < 0 \rightarrow \begin{cases} \text{colder box 1} \\ \text{warmer box 2} \\ \text{slightly colder box 3} \end{cases}$$

$$\rightarrow \begin{cases} \text{increase in NH atmospheric transports} \\ \text{weaker increase in SH atmospheric transports.} \end{cases}$$

Both atmospheric transports advect buoyancy from the equator to the poles. Determining the sign of the EMT feedback is straightforward. For a negative flow perturbation, more anomalous moisture is advected in the Northern Hemisphere than the Southern Hemisphere, further weakening flow, that is producing a positive

feedback loop. By increasing the freshening of boxes 1 and 3, the EMT feedback directly impedes the system's ability to redistribute anomalous equatorial salinity. A numerical determination of the eigenvalues confirms this result, as plotted in Fig. 8. If the EMT feedback is turned on (changing $m = 0$ to $m = 3$), the system is always less stable, particularly in case III where the EMT feedback works in conjunction with the positive ocean salinity feedback.

The effect of the EHT feedback, on the other hand, is more complicated. Figure 8 illustrates the growth rate as n is varied, including the limit of fixed temperature gradients (i.e., for $n \rightarrow \infty$; note that the EMT feedback is now inoperative). Similar to the EMT feedback, the EHT feedback advects more anomalous buoyancy to the sinking hemisphere following a negative flow perturbation, producing a positive feedback loop. To understand the behavior in Fig. 8, one must also examine the interaction between the EHT feedback and the model's other feedbacks. As mentioned, increasing n limits the model's temperature deviations, which in turn limits the effectiveness of the positive EMT feedback. For example, note in case III that the growth rate decreases as the EHT feedback is turned "on" (i.e., n changes from 0 to 1) only if the EMT feedback is active. By restoring temperature more rapidly, the EHT also limits the effectiveness of the negative ocean temperature feedback. These two interactions also occur in the single-hemisphere model, as described in M96; changes in the ocean salinity feedback among the three cases complicates the analysis here.

The interaction between the EHT and the ocean temperature feedback is particularly important in case III, as the positive ocean salinity feedback implies that the negative feedbacks must work harder to restore the equilibrium salinity structure. Hence, the model is extra sensitive to any process that hinders the effectiveness of the ocean temperature feedback. Conversely, in case II
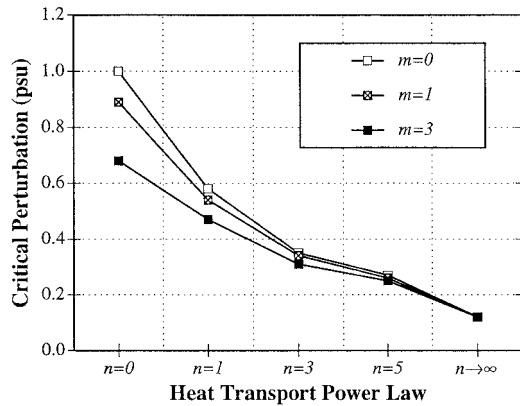
Fig. 9. Critical perturbation necessary to induce permanent transition to southern sinking given varying atmospheric heat and moisture flux power laws.

Table 5. Critical perturbation found when the eddy transport processes are prescribed independently for each hemisphere. For both heat and moisture transport, OFF represents fixed fluxes whereas a linear power law is employed for the ON state.

		SH feedbacks	
		OFF	ON
NH feedbacks	OFF	−1.00	−1.20
	ON	−0.46	−0.54

and nonlinear stability behavior was consistent (M96; Krasovskiy and Stone 1998).

To explain the observed destabilization for larger n, we first recall from section 4 that a sharp rebound in overturning strength occurs following a large perturbation due to a strong ocean temperature feedback. In the limit as $n \to \infty$ this rebound vanishes, and the equilibrium is less stable, despite the fact that the destabilizing EMT feedback is effectively shut off. This suggests that the magnitude of the temperature rebound plays the largest role in nonlinear stability, whereas the accompanying change in the EMT feedback strength as n is varied (by limiting temperature deviations) is secondary.

In section 4 it is noted that the flow strength of the quasi-no-flow solution depends upon asymmetries in freshwater forcing. By augmenting the Northern Hemisphere moisture flux following a critical perturbation (recall that T_N subsequently increases), the EMT feedback effectively places the intermediate solution closer to the starting equilibrium in phase space. Thus, a smaller critical perturbation is necessary as m is increased. Not surprisingly, the decrease in nonlinear stability as n is increased is most similar to the linear stability behavior for case III, as we have argued the stability criteria converge as the neutrally stable equilibrium is approached.

To isolate the importance of the Southern versus the Northern Hemisphere atmospheric feedbacks, we now prescribe the eddy transport power laws independently. Critical perturbation results are presented in Table 5. The Southern Hemisphere feedbacks, in isolation, are weakly stabilizing whereas the Northern Hemisphere feedbacks are strongly destabilizing. The linear stability behavior is similar (an exception to this is given by case II, where the behavior observed in Fig. 8 leads to weak destabilization as n increases in the Southern Hemisphere). Given that the model has a weak southward oceanic heat flux in the Southern Hemisphere, when in reality the South Atlantic has a modest northward flux (e.g., Macdonald and Wunsch 1996), the stabilization provided by the Southern Hemisphere feedbacks is likely an artifice of the model's simplicity. In other words, our ''backwards'' heat flux in the South Atlantic reverses the sign of the Southern Hemisphere feedback; with a more faithful representation, the South Atlantic is heated following a negative flow anomaly, producing a destabilizing feedback. The point we wish to stress,

the ocean salinity feedback is negative, providing an alternative means for restoring salinity balance. In other words, the ocean temperature feedback is less important, and consequently restoring temperature more rapidly via the EHT feedback provides an additional ''boost'' in restoring equilibrium conditions (also note that the presence of the negative salinity feedback lessens any destabilization due to the EMT feedback). For case I, a trade-off occurs: Increasing n restores temperatures more rapidly, but also weakens the ocean temperature feedback, so no significant change in the growth rate is observed.

Our strategy in this section has been to examine stability behavior while preserving the mean overturning strength through our choice of χ_n and γ_n. If we instead examine how stability depends upon temperature damping (either radiative or through the atmospheric heat fluxes) while permitting the mean flow to change, we find that stronger restoring stabilizes by increasing the mean flow feedback (this is also true in Stommel's model for the stable northern sinking equilibria; see MS). In other words, changes in the mean flow overwhelm any changes in the advective feedbacks. In OGCMs, however, stronger restoring destabilizes, as changes in a thermal convective feedback (not present in Rooth's model) apparently overwhelm changes in the mean flow feedback (Zhang et al. 1993; Rahmstorf and Willebrand 1995).

To extend the results of section 4, we determine the critical perturbations for various m and n using the reformulated model. Results are illustrated for symmetric freshwater forcing in Fig. 9. Although case II is in general more stable and case III is less stable, the same qualitative behavior emerges: Stability always decreases if one increases either exponent. This behavior is different from the linear stability analysis, where the response depends upon the freshwater forcing scenario. This result also runs counter to that of the single-hemisphere models, where it was observed that the linear

however, is that the Northern Hemisphere eddy moisture and heat flux transports are dominant while the Southern Hemisphere feedbacks play a less important role due to the Southern Hemisphere's weaker oceanic heat flux.

The dominance of the Northern Hemisphere feedbacks in the model is interesting in light of the results in section 2, where we determine that the overturning strength is governed by the freshwater forcing in the Southern Hemisphere. Using a coupled ocean–atmosphere general circulation model, Wang et al. (1999b) suggests both feedbacks are in fact destabilizing, consistent with a more reasonable oceanic heat transport in the South Atlantic, although it is not clear if the Northern Hemisphere feedback is any stronger than the Southern Hemisphere feedback. Furthermore, combining the feedbacks does not seem to reduce stability further. Despite limitations in our model that prevent us from addressing these more interesting behaviors, our results suggest that previous studies employing the single-hemisphere model (e.g., Nakamura et al. 1994 and M96) are largely applicable in the interhemispheric context.

To test the robustness of results presented in this section, the mean (reference) state is re-equilibrated so that the atmospheric heat and moisture transport efficiencies are symmetric about the equator for $m = n = 3$, reflecting nonlinear atmospheric transport processes. The resulting equilibria are modestly more stable than for the $m = n = 1$ mean state, but similar qualitative behavior for different EMT and EHT feedback strengths is observed.

6. Summary and discussion

Using the understanding that has emerged from studies based on Stommel's (1961) intrahemispheric model as a starting point, our effort here is to forge a similar conceptual picture for the interhemispheric circulation. Our focus is on understanding the stability behavior, both in ocean-only and coupled ocean–atmosphere modes, using a box model of a single ocean basin where overturning strength is parameterized as a function of the density difference between two high-latitude boxes, as first proposed by Rooth (1982). This work explicitly considers the implications of asymmetric freshwater forcing in the two hemispheres, which is beyond the scope of the single-hemisphere models and has received scant attention in the literature [Rahmstorf (1996) is a notable exception], perhaps because freshwater forcing in the high-latitude South Atlantic has not been determined to high precision (Broecker et al. 1990).

Although freshwater forcing in the Southern Hemisphere alone governs overturning strength, which was previously found by Rahmstorf (1996), we find that increasing freshwater forcing in the Northern Hemisphere leads to a heretofore unrecognized instability in the northern sinking branch. More generally, the ratio of Northern Hemisphere over Southern Hemisphere freshwater fluxes emerges as the most important parameter

governing the stability of the northern sinking solution. These results are in remarkable accord with Wang et al. (1999a); in a global OGCM of idealized geometry, they concluded that NADW production was largely governed by the Southern Hemisphere freshwater flux, but stronger moisture fluxes in the Northern Hemsiphere weakened the stability of the overturning circulation. In our model, this behavior is attributed to the ocean salinity feedback, which becomes positive once the northern freshwater forcing is made larger than the southern forcing [this is the case for the actual estimated freshwater forcing, as given by Broecker et al. (1990) based on Baumgartner and Reichel's (1975) study]. If the northern forcing is instead made weaker than the southern forcing, the feedback becomes increasingly negative, and the equilibrium subsequently more stable. In this manner Rooth's model is quite different from Stommel's, where this feedback is always positive. The ocean temperature feedback, the mean flow, and Newtonian cooling, on the other hand, are found to operate similarly to respective processes in the single-hemisphere model.

Our results here contrast with those of Rahmstorf (1996), who suggested that the northern sinking branch is driven by north to south temperature variations, ending as a saddle-node bifurcation for increased (equatorward) atmospheric moisture transport in the Southern Hemisphere. Here we argue that the overturning circulation in the Atlantic is instead driven by salinity variations and that the equilibrium becomes unstable at a subcritical Hopf bifurcation as freshwater forcing in the Northern Hemisphere is increased. Equatorward atmospheric moisture flux is not consistent with observations (Peixoto and Oort 1992); moreover, the multiple equilibria in the global coupled GCM of Manabe and Stouffer (1988) are clearly determined by salinity effects.

If the equilibrium state is perturbed sufficiently, the model is attracted to a transient, unstable intermediate solution. This "quasi-no-flow" solution is characterized by weak flow unless the north–south asymmetry in freshwater forcing is such that the equilibrium branch approaches the Hopf bifurcation, into which the intermediate state converges. For symmetric freshwater forcing, the quasi-no-flow solution reduces to the no-flow solution discussed by Rooth (1982). The equilibrium thermal structure (i.e., the northern box is warmer than the southern box) resists a permanent transition, producing a strong rebound in flow strength following the perturbation. The quasi-no-flow solution has effectively lost all memory of its initial equilibrium state, hence a critical perturbation is just large enough to place the system almost exactly in this intermediate state. Since the quasi-no-flow solution is more easily attained (via salt perturbations) for increasing flow strength of the intermediate solution, stronger freshwater forcing in the Northern Hemisphere leads to nonlinear destabilization and vice versa, consistent with the linear stability results.

Due to the asymmetry in oceanic heat flux between

hemispheres, the atmospheric feedbacks in the sinking hemisphere are more important for linear stability behavior. In this context, we conclude that analyses based on the Stommel model are largely applicable due to the atmospheric processes' inherent dependence on the equator-to-pole temperature gradient. The eddy moisture transport feedback is always positive, but increasing the eddy heat transport's power law relation may be stabilizing or destabilizing, depending primarily on the sign of the ocean salinity feedback. An atmosphere that efficiently damps changes in its temperature structure can moderate the effect of a strong moisture feedback [as discussed in Marotzke (1996)], providing stabilization. On the other hand, this temperature damping also moderates the effect of the stabilizing ocean temperature feedback. The ocean salinity feedback tips the balance of this trade-off; given a positive salinity feedback, this latter effect leads to destabilization, whereas the former dominates the behavior given a negative salinity feedback. Unlike the linear stability behavior, more efficient temperature damping by the eddy heat transport always leads to reduced finite amplitude stability by limiting the ''inertia'' provided by the equilibrium temperature structure.

Our results offer a possible explanation for why Mikolajewicz and Maier-Reimer (1994) saw an increased stability of the ''conveyor belt'' in their OGCM under weaker Newtonian damping, whereas ''southern sinking'' became less stable. Notice, first, that without external asymmetries, the ''catchment areas'' of the northern and southern sinking equilibria must be equally large and independent of the Newtonian damping coefficient λ. However, the ''demarcation'' lines in phase space can and do change, so the evolution of any particular initial state depends on λ. Apart from the asymmetric topography in Mikolajewicz and Maier-Reimer's global model, they restored surface temperatures to an asymmetric field with the conveyor belt solution ''built-in.'' We stress here that there is nothing inherent in either weak or strong thermal ocean–atmosphere coupling that would favor either equilibrium.

In coupled ocean–atmosphere models, a state with minimal Atlantic deep-water formation similar to our quasi-no-flow solution was observed by Manabe and Stouffer (1988) and Wang et al. (1999a). It is not clear if either model would have spontaneously switched to a conveyor-type circulation had they been integrated further. More likely, the ''drop dead'' state would have undergone periodic flushes, similar to those observed in Marotzke (1989) and Weaver and Sarachik (1991), which are impossible in Rooth's model due to the lack of an explicit convection parameterization. Here, the model is forced to leave the ''drop dead'' state, destined for either the northern or southern sinking equilibrium branch. The requirement that the critical perturbation must immediately reverse circulation for all but the least stable equilibria also calls into question the limitations of the conceptual model.

We summarize as follows. Our model, albeit simple, provides a natural explanation for how poleward moisture transport in the Southern Hemisphere dominates interhemispheric thermohaline flow, while Northern Hemisphere moisture flux is crucial for the circulation's stability. North–south temperature asymmetries have also been shown to be important for stability, and feedbacks with the atmosphere demonstrated to be dominated by Northern Hemisphere effects.

REFERENCES

Baumgartner, A., and E. Reichel, 1975: *The World Water Balance.* Elsevier, 179 pp.

Broecker, W. S., T.-P. Peng, J. Houzel, and G. Russell, 1990: The magnitude of global freshwater transports of importance to ocean circulation. *Climate Dyn.,* **4,** 73–79.

Bryan, F., 1986: High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature,* **323,** 301–304.

Gordon, A. L., 1986: Interocean exchange of thermocline water. *J. Geophys. Res.,* **91,** 5037–5046.

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

Klinger, B. A., and J. Marotzke, 1999: Behavior of double-hemisphere thermohaline flows in a single basin. *J. Phys. Oceanogr.,* **29,** 382–399.

Krasovskiy, Y., and P. H Stone, 1998: Destabilization of the thermohaline circulation by atmospheric transports: An analytic solution. *J. Climate,* **11,** 1803–1811.

Levitus, S., 1982: *Climatological Atlas of the World Ocean.* NOAA Prof. Paper No. 13, U.S. Dept. of Commerce, NOAA, Washington, D.C., 173 pp.

Lohmann, G., R. Gerdes, and D. Chen: 1996. Stability of the thermohaline circulation in a simple coupled model. *Tellus,* **48A,** 465–476.

Macdonald, A. M., and C. Wunsch, 1996: An estimate of global ocean circulation and heat fluxes. *Nature,* **382,** 436–439.

Manabe, S., and R. J. Stouffer, 1988: Two stable equilibria of a coupled ocean–atmosphere model. *J. Climate,* **1,** 841–866.

——, and ——, 1995: Simulation of abrupt climate change induced by freshwater input to the North Atlantic Ocean. *Nature,* **378,** 165–167.

Marotzke, J., 1989: Instabilities and multiple steady states of the thermohaline circulation. *Ocean Circulation Models: Combining Data and Dynamics.* D. L. T. Anderson and J. Willebrand, Eds., Kluwer, 501–511.

——, 1990: Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. thesis, Ber. Inst. Meeresk., Kiel, Germany, 126 pp.

——, 1996: Analysis of thermohaline feedbacks. *Decadal Climate Variability: Dynamics and Predictability.* D. L. T. Anderson and J. Willebrand, Eds., Springer-Verlag, 333–378.

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

——, 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, 1994: Mixed boundary conditions in ocean general circulation models and their influence on the stability of the model's conveyor belt. *J. Geophys. Res.,* **99,** 22 633–22 644.

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

Peixoto, J. P., and A. H. Oort: 1992. *Physics of Climate.* Amer. Inst. Phys., 520 pp.

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

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

——, J. Marotzke, and J. Willebrand, 1996: Stability of the thermohaline circulation. *The Warm Water Sphere of the North Atlantic Ocean,* W. Krauss, Ed., Borntraeger, 129–157.

Rooth, C., 1982: Hydrology and ocean circulation. *Progress in Oceanography,* Vol. 11, Pergamon, 131–149.

Saravan, R., and J. C. McWilliams, 1995: Multiple equilibria, natural variability, and climate transitions in an idealized ocean–atmosphere model. *J. Climate,* **8,** 2296–2323.

Schiller, A., U. Mikolajewicz, and R. Voss, 1997: The stability of the thermohaline circulation in a coupled ocean–atmosphere general circulation model. *Climate Dyn.,* **13,** 325–347.

Stocker, T. F., and D. G. Wright, 1991: Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature,* **351,** 729–732.

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

Stone, P. H., and D. A. Miller, 1980: Empirical relations between seasonal changes in meridional temperature gradients and meridional fluxes of heat. *J. Atmos. Sci.,* **37,** 1708–1721.

——, and M.-S. Yao, 1990: Development of a two-dimensional zonally averaged statistical–dynamical model. Part III: The parameterization of the eddy fluxes of heat and moisture. *J. Climate,* **3,** 726–740.

Tziperman, E., 1997: Inherently unstable climate behavior due to weak thermohaline ocean circulation. *Nature,* **386,** 592–595.

Wang, X., P. H. Stone, and J. Marotzke, 1999a: Global thermohaline circulation. Part I: Sensivity to atmospheric moisture transport. *J. Climate,* **12,** 71–82.

——, ——, and ——, 1999b: Global thermohaline circulation. Part II: Sensivity with interactive atmospheric transports. *J. Climate,* **12,** 83–91.

Warren, B. A., 1993: Why is no deep water formed in the North Pacific? *J. Mar. Res.,* **41,** 327–347.

Weaver, A. J., and E. S. Sarachik, 1991: The role of mixed boundary conditions in numerical models of the ocean's climate. *J. Phys. Oceanogr.,* **21,** 1470–1493.

——, J. Marotzke, P. F. Cummins, and E. S. Sarachik, 1993: Stability and variability of the thermohaline circulation. *J. Phys. Oceanogr.,* **23,** 39–60.

Welander, P., 1986: Thermohaline effects in the ocean circulation and related simple models. *Large-Scale Transport Processes in Oceans and Atmosphere,* J. Willebrand and D. L. T. Anderson, Eds., NATO ASI Series, D. Reidel, 163–200.

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