Multiple Equilibrium States in Combined Thermal and Saline Circulation*

RUI XIN HUANG, JAMES R. LUYTEN, AND HENRY M. STOMMEL

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 10 August 1990, in final form 8 July 1991)

ABSTRACT

Structure and stability of the multiple equilibria of the thermohaline circulation are studied using 2 × 2 and 3 × 2 box models. Thermohaline catastrophe is a shallow phenomenon and its time evolution consists of three stages: the search stage, the catastrophic stage, and the adjustment stage.

A 3 × 2 box model is introduced to permit fitting a somewhat less truncated form of freshwater forcing. Although the total number of possible modes of circulation increases quickly as the number of boxes is increased, the grand thermal mode and grand saline mode remain. Some multiple solutions are found to be statistically improbable, and they belong to the so-called minor modes. With the more realistic freshwater forcing appropriate for the present-day North Atlantic, our simple 3 × 2 model predicts that the thermohaline circulation is near a critical state, because a small increase in freshwater flux can cause the system to collapse from the present thermally dominated circulation to an intermediate mode of circulation with strong sinking at midlatitude and a rather sluggish circulation in the polar basin.

1. Introduction

Gravitational convection due to combined thermal and haline forcing in an upwind hydrostatic box model with linear frictional dynamics exhibits multiple equilibrium states. In applying such models to the ocean, the so-called Haney (1971) boundary condition has been widely used for both temperature and salinity fields. In fact, the external forcing of these two fields involves quite different physical processes and should be parameterized in quite different ways. The effect of different boundary conditions on temperature and salinity was discussed by Stommel (1961), who showed that two stable states can arise because of the difference in the Rayleigh coefficient for the two constituents. Rooth (1982) investigated the case of a three-box model, consisting of an equatorial box and two polar boxes, and suggested that a symmetric solution might be unstable and could drift toward a pole–pole mode. Welander (1986) pointed out that such a three-box model can have four stable equilibria: two symmetric and two asymmetric.

Rooth's speculation was confirmed by Bryan's (1986) numerical experiments based on the GFDL primitive equation model, in which a freshwater flux is specified as the upper boundary condition for the salinity field. Bryan found that the solution, which is symmetric with respect to the equator, is indeed unstable to finite-amplitude perturbation in salinity, under which the so-called halocline catastrophe takes place and the system drifts toward one of the stable pole–pole modes.

The instabilities and multiple steady states associated with the subtle difference in boundary conditions for the temperature and salinity have been further discussed, using simple box models and a numerical model on a meridional section, by Marotzke et al. (1988) and Marotzke (1989). A double-diffusion instability associated with these different boundary conditions for the temperature and salinity fields was discussed by Welander (1989).

Marotzke (1990) made extensive studies of the multiplicity of thermohaline circulations, using a multibox model for the World Ocean and two- and three-dimensional primitive equation numerical models.

The concept of multiple equilibria and halocline catastrophe of the thermohaline circulation has been applied to the Younger Dryas period for a possible explanation by Maier-Reimer and Mikolajewicz (1989).

Since the oceanic component of the ocean–atmosphere coupled system can have at least two stable states, it seems natural to inquire about the bimodality of the coupled system. The possibility that the ocean–atmosphere coupled system can have two stable states was first explored by Broecker et al. (1985). The first example of the two stable modes in the ocean–atmosphere coupled system was found in a coupled GCM model by Manabe and Stouffer (1988). Using a Five-

* Contribution No. 7502 from the Woods Hole Oceanographic Institution.

Corresponding author address: Dr. Rui Xin Huang, Woods Hole Oceanographic Institute, Department of Physical Oceanography, Woods Hole, MA 02543.

© 1992 American Meteorological Society

box model for a single hemisphere, the bimodality of the ocean–atmosphere coupled system was studied by Birchfield (1989) and Birchfield et al. (1990).

As a nonlinear system, the thermohaline circulation possesses multiple solutions. Two types of solution, the grand thermal mode (sinking at the pole) and the grand saline mode (upwelling at the pole), have received the most attention. However, the thermohaline circulation in the oceans can have many solutions involving minor modes.

There is apparently little paleoceanographic evidence that the grand saline mode occurred in the North Atlantic during recent geological history, although it might have existed during the Cretaceous period, with a very different (from the present-day) ocean–atmosphere system. During the Cretaceous, the climate was uniformly warm and wet. Warm and salty water masses might have been formed due to intensive evaporation in the shallow coastal seas and sank to the bottom, then upwelled somewhere else. The oceanic circulation system at that time is yet unclear; for climate during the Cretaceous, see the review by Barron (1983).

Paleoceanographic studies for the last glaciation show that nutrient concentrations were higher in the deeper waters of the glacial North Atlantic than those today, indicating that deep flow from nutrient-depleted source waters was less then. However, that these concentrations remained lower than those of the Pacific suggests that flow patterns were not reversed during glacial times (see the review by Boyle 1990).

As pointed out by Duplessy et al. (1988), during the last glacial maximum, the deep-water circulation pattern was noticeably different from the present. The southern source of deep water was the major feature of global deep-water circulation, although a relatively small amount of bottom water was also formed in the North Atlantic. There is no evidence that the thermohaline circulation was totally reversed during that period. On the other hand, the amount of intermediate water was substantially enhanced. That is, the deep-water formation was slowed down, but not totally cut off, while this decline in deep-water formation was compensated with an enhancement in intermediate water formation.

Paleoceanographic studies also suggest that the deep circulation in the North Atlantic exhibits many different modes (Oppo and Fairbanks 1990) or a continuum of circulation states (Raymo et al., 1990), instead of simple bimodality. The differences among some of these different states of thermohaline circulation are not as dramatic as those involved with a total reversal of the meridional circulation; nevertheless, they represent some profound changes in the oceanic circulation and water mass formation.

Box models illustrate the effects of nonlinear kinematic processes (such as the advection of buoyancy) in conceptually simple ways over a wide range of parameters. They cannot be expected to approach the fine-scale resolution and more correct and complete dynamics of sophisticated finely resolved gridded general circulation models. Box models do not compete with numerical models.

A modest increase in the number of boxes can be useful to address two issues: 1) to test for sensitivity to truncation and 2) to detect inexpensively preferred spatial modes or symmetries of the system by large numbers of runs with a variety of initial conditions. Exploring what determines the location and size of sinking regions, relative volume fluxes of surface, intermediate and bottom waters, and the asymmetry of the meridional cell suggests that some phenomenon, such as sinking in a small region, in the large numerical models (and nature) are primarily due to the nonlinearity associated with density advection. Accordingly, elementary interpretations can be inferred from simple analysis of the physical processes involved.

In this study we will first formulate a 2×2 box model, with a Rayleigh boundary condition for temperature and a flux boundary condition for salinity. When the amplitude of freshwater flux passes through critical values, catastrophic transitions from one state to another occur. These critical values, and the time required to pass through the catastrophe from one equilibrium to another, are described. To show the size of the basin of attraction for the thermal mode and the saline mode, we use a Monte Carlo method to calculate the possibility partition as a function of the strength of the precipitation. A similar approach was used by Marotzke (1990) to discuss the relative frequency distribution of different modes.

A 2×2 box model can at best represent only the horizontally grand modes of forcing, and a single type of deep water. A 3×2 box is introduced to permit fitting a reasonably realistic minor modal form (in latitude) of the freshwater flux. As soon as the number of boxes in the model is increased to 3×2, the number of possible modes is increased to 6. However, these modes can be classified into the grand thermal mode and saline mode and some minor, or intermediate, modes. Most importantly, although the total number of possible modes increases, these modes are not equally important. By using a Monte Carlo method, we will show that some modes are dominant while some minor modes are very unlikely to appear.

With a simple 3×2 box model, we discuss several cases. One of these cases is forced with a "realistic" freshwater flux, and the model suggests that for the freshwater flux close to the present day, the thermohaline circulation in the North Atlantic is very close to a critical point where a small perturbation can cause the system to fall from a thermal state with strong sinking at high latitude and weak sinking at midlatitude to an intermediate mode with strong sinking at midlatitude and weak upwelling at high latitude. A 3×2 box model does not have enough vertical resolution to differentiate between deep and intermediate depths, so

geometrically we cannot speak of deep and intermediate waters. However, in the ocean, deep water generally forms at the poleward limits of an ocean basin, and intermediate water at subpolar latitudes. With this caution, it seems permissible to associate the idea of intermediate water formation as a phenomenon with minor (or intermediate) modes, where sinking in the middle box of a single-basin model might be thought of as the model's attempt to form intermediate water despite the lack of an intermediate depth at which it could level off.

With appropriate rescaling the 3 × 2 model is also used to study the case of circulation patterns in pole-to-pole oceans. The forcing is chosen to be symmetrical about the equator. For realistic amplitude of imposed freshwater flux the system is most likely to be in a pole-to-pole mode in which one hemisphere has a weak saline circulation and the other a strong thermal mode of circulation, either polarity being equally likely. Of course, the preference for pole-to-pole modes diminishes the relevance of the single-basin model that is the main subject of this paper.

2. The 2 × 2 box model

a. *The model formulation*

We begin with a 2 × 2 box model, as shown in Fig. 1. Note that the lower-layer thickness is δh, where $0 \leqslant \delta \leqslant \infty$. The reference fixed air temperature is $T_1^* = T^*$ and $T_2^* = 0$. Within each box, water properties are well mixed and the density is calculated using a linear state equation

$$\rho_i = \rho_0(1 - \alpha T_i + \beta S_i), \quad i = 1, 2, 3, 4. \tag{1}$$

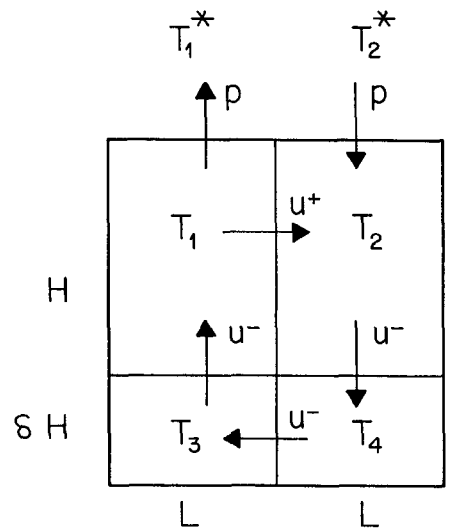

FIG. 1. The 2 × 2 box model defining applied air temperature T_i^*, amplitude of precipitation p, temperatures of boxes T_i, velocities u^+, u^-, box thicknesses H and δH, and box width L.

The pressure is calculated by the hydrostatic relation. In the horizontal direction velocity is linearly proportional to the pressure gradient ∇P,

$$u^+ = -c\frac{P_1 - P_2}{L}, \quad u^- = c\frac{P_3 - P_4}{L}. \tag{2}$$

Note that rotation is neglected because it is almost impossible to introduce rotation into such a highly truncated box model. The upper boundary condition is a Rayleigh condition for the temperature; that is, the heat flux is

$$H_f = \rho_0 c_p \Gamma(T^* - T) \tag{3}$$

where Γ is a Rayleigh relaxation coefficient with T^* prescribed, and a flux boundary condition for the salinity. Freshwater flux is specified at each upper-layer box. The algebraic sum of the applied freshwater fluxes must be zero for a steady model.

The upwind scheme is used for the advection terms in the heat and salinity equations. Assuming the velocity pattern is that in Fig. 1, we can write down the heat and salinity balance for each box. We introduce the nondimensional variables

$$T_i = T_0 T_i', \quad S_i = S_0 S_i'$$

$$(u^+, u^-) = \frac{L}{H}\frac{\Gamma}{\rho_0 c_p}(u^{+\prime}, u^{-\prime}), \quad w_i = \frac{\Gamma}{\rho_0 c_p} w_i'$$

$$p = \frac{\Gamma}{\rho_0 c_p} p', \quad t = \frac{H\rho_0 c_p}{\Gamma} t', \tag{4}$$

where $T_0 = 25°C$ is the imposed north–south temperature difference, $T_1^* = T^* = T_0$, $T_2^* = 0$, $S_0 = 35$ psu (practical salinity units) is the mean salinity, $H = 2$ km, $L = 3000$ km, w_i is the vertical velocity, and p is the precipitation rate. After dropping the primes, the nondimensional equations for heat and salinity balance in the upper two boxes are

$$\frac{dT_1}{dt} = -u^+ T_1 + u^- T_3 + 1 - T_1$$

$$\frac{dT_2}{dt} = u^+ T_1 - u^- T_2 - T_2$$

$$\frac{dS_1}{dt} = -u^+ S_1 + u^- S_3$$

$$\frac{dS_2}{dt} = u^+ S_1 - u^- S_2. \tag{5}$$

The equations for the lower two boxes are

$$\frac{dT_3}{dt} = \frac{u^-}{\delta}(T_4 - T_3)$$

$$\frac{dT_4}{dt} = \frac{u^-}{\delta}(T_2 - T_4)$$

$$\frac{dS_3}{dt} = \frac{u^-}{\delta}(S_4 - S_3)$$

$$\frac{dS_4}{dt} = \frac{u^-}{\delta}(S_2 - S_4) \tag{6}$$

where the velocity can be calculated by

$$u^+ = -\frac{p}{2} + A(T_1 - T_2 + \delta T_3 - \delta T_4) - B(S_1 - S_2 + \delta S_3 - \delta S_4)$$

$$u^- = u^+ + p \tag{7}$$

where the two constants A and B are defined as

$$A = C\alpha T_0$$

$$B = C\beta S_0$$

$$C = \frac{cgH^2\rho_0{}^2 c_p}{4\Gamma L^2}.$$

According to Haney (1971), $\Gamma/(\rho_0 c_p) = 0.7$ m day^{-1} $= 8.1 \times 10^{-4}$ cm s^{-1}, so that the time scale is

$$\rho_0 c_p \frac{H}{\Gamma} = \frac{2000}{0.7} \text{ days} = 2857 \text{ days} \approx 8 \text{ years}.$$

In the system there are four nondimensional parameters: A, B, δ, and p.

How can we determine C? In the oceans the large-scale vertical velocity w is no more than 10^{-4} cm s^{-1}. Thus, the nondimensional w is

$$w = \frac{10^{-4} \text{ cm s}^{-1}}{0.7 \text{ m day}^{-1}} \approx 0.12.$$

By running the model, we have found that a plausible estimate for C, which gives rise to appropriate values of w, is about 0.05. The corresponding p for the present-day North Atlantic is about 0.38×10^{-5} cm s^{-1}.

Since an upwind scheme is used in the model, the equation set, (5) and (6), is valid only for a case when both u^+ and u^- are positive; otherwise, the convection terms should be changed to maintain upwind conditions.

An overturning scheme is implemented in the model so that whenever a gravitationally unstable situation appears within a vertical column, water properties will be mixed vertically to eliminate the unstable stratification. This scheme is called a convective adjustment in many numerical models. Since convective motion is induced by any horizontal motion in this model, we will call the scheme "overturning."

Apart from mixing associated with advection and overturning, there is no other type of mixing in the model, such as turbulent mixing between boxes or double-diffusive mixing. As a result of the mixing parameterization, uniformity of the bottom boxes is a result of sinking from a single surface box. In a more elaborate model, the uniformity of the bottom boxes can be changed by overturning induced by wind forcing or cross-box turbulent mixing; however, we will not discuss these cases in order to make the model more focused. In a separate study we have found that double diffusion (in a parameterized form) does not qualitatively change the bifurcation nature of the model.

b. Reduction to a 2-box model

An interesting case is the limit when $\delta \to 0$; then the model is reduced to a 2×1 box model. Essentially, the lower layer degenerates into a pipe; its sole function is returning the water back to box 1 so as to close the circulation. In this limit, and assuming no overturning in the southern column, the heat and salinity response time in boxes 3 and 4 is zero, so that when $u^+ > 0$

$$T_3 = T_4 = T_2, \quad S_3 = S_4 = S_2.$$

Therefore, the equations are reduced to

$$\frac{dT_1}{dt} = -u^+T_1 + u^-T_2 + 1 - T_1$$

$$\frac{dT_2}{dt} = u^+T_1 - u^-T_2 - T_2$$

$$\frac{dS_1}{dt} = -u^+S_1 + u^-S_2$$

$$\frac{dS_2}{dt} = u^+S_1 - u^-S_2. \tag{8}$$

The velocities are determined by

$$u^+ = -\frac{p}{2} + C(\mathcal{T} - \mathcal{S}) \tag{9}$$

and

$$u^- = u^+ + p \tag{10}$$

where

$$\mathcal{T} = \alpha T_0(T_1 - T_2), \quad \mathcal{S} = \beta S_0(S_1 - S_2).$$

We distinguish three regions in $\mathcal{S}$, $\mathcal{T}$ phase space: region I where $u^+ \geq 0$, $u^- \geq 0$; region II where $u^+ < 0$, $u^- \leq 0$; and region III where $u^+u^- < 0$.

If we now look for steady equilibrium solutions of each of these equations, we set $dT_i/dt = 0$, $dS_i/dt = 0$ and write (using $\mathcal{S}_0 = \beta S_0$, $\mathcal{T}_0 = \alpha T_0$)

Heat equilibrium curve on $\mathcal{S}$, $\mathcal{T}$ plane *Region*

$$\mathcal{S} = \frac{1}{2C} + \mathcal{T} - \frac{(1+p)}{2C}\frac{\mathcal{T}_0}{\mathcal{T}} \qquad \text{I}$$

$$\mathcal{S} = -\frac{1}{2C} + \mathcal{T} + \frac{(1+p)}{2C}\frac{\mathcal{T}_0}{\mathcal{T}} \qquad \text{II} \tag{11}$$

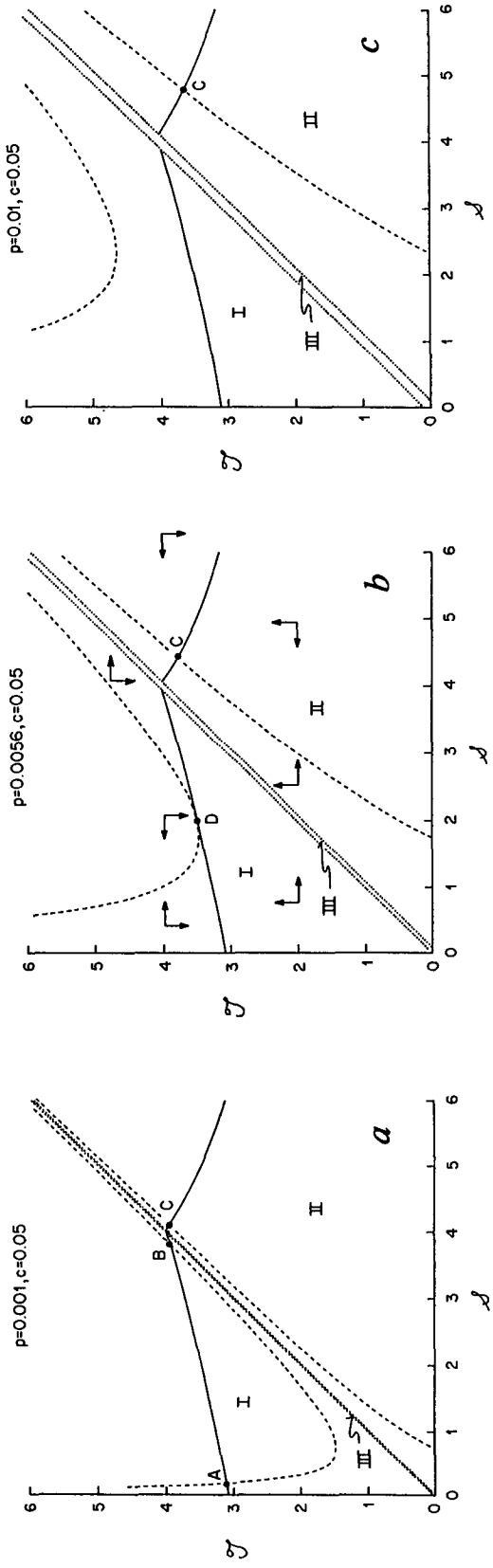

FIG. 2. $\mathcal{S}$–$\mathcal{T}$ phase diagram for the 2×1 box ($\delta \to 0$). Three panels are for different values of p (nondimensional). The phase space consists of three regions, I, II, and III, characterizing the velocity pattern. The curve for heat equilibrium is solid; that for salt equilibrium is dashed. Intersections of these two sets of curves define equilibrium states. The intersection at point A is a stable thermally driven equilibrium. Point B is unstable. Point C is a stable salinity-dominated equilibrium. Point D represents the coalescence of points A and B at the critical $p_c = 0.0056$. In panel (b) there are pairs of arrows to show the directions of $\mathcal{S}_t$ and $\mathcal{T}_t$ in regions bounded by the equilibrium curves.

Salt equilibrium curve on $\mathcal{S}$, $\mathcal{T}$ plane	*Region*
$\mathcal{T} = \mathcal{S} + \frac{p}{C}\frac{\mathcal{S}_0}{\mathcal{S}}$	I
$\mathcal{T} = \mathcal{S} - \frac{p}{C}\frac{\mathcal{S}_0}{\mathcal{S}}$	II. (12)

There are no consistent equilibrium curves for region III on the $\mathcal{S}$, $\mathcal{T}$ plane. The regions are mapped on the $\mathcal{S}$, $\mathcal{T}$ plane in Fig. 2a. The unsteady region III lies astride the diagonal $\mathcal{T} = \mathcal{S}$ where $u^+ < 0$ and $u^- > 0$, and hence, both velocities flow from left to right (very slowly, of course), tending to sweep all the salt out of box 2 into box 1.

The equilibrium curves are shown in Figs. 2a–c for the canonical choice of constants, with the precipitation p treated as a parameter. The region III is very narrow because over most of the plane the range of p chosen is very small compared to the velocities u^+ and u^- in the circulation: p is an order of magnitude less than u^+ or u^-.

For $p = 0.001$ (panel a) the curves have three intersections, so these define equilibrium states for the system. The point A in region I is a stable node with a thermally dominated circulation. The point B is an unstable node in region I; it repels the phase point. Point C is a stable spiral in region II with a salinity-dominated circulation. The sense of circulation for the equilibrium states at points A and C is opposed.

As p is increased to $p = 0.0056$ (panel b), the salt curve tends to pull away from the temperature curve in region I, so that points A and B coalesce to point D, which defines the critical value of $p = p_c$ where the two curves are tangent. This is the critical situation where "catastrophe" occurs. Values of p_c are tabulated in Table 1, as functions of δ and C. It is seen that p_c is a weak function of δ, indicating that the catastrophe is essentially a surface phenomenon, independent of the bottom water. This aspect of the physics will also emerge from a study of the time evolution of the system through p_c. On the other hand, p_c is a strong function of C, the "conductance" of the model. Large values of C correspond to large mass fluxes u^+ for a given horizontal pressure gradient. Another way to think of the physical role of C is to remember that it is proportional to gravity. Consequently p_c increases with C. If p is increased further, say to $p = 0.01$ (panel c), the two curves in region I no longer intersect and there are no equilibrium states in the region. The only steady state that exists is the salinity-dominated one at point C. Therefore, the phase point moves from D to C, passing through region III into region II.

It is easy to visualize how a phase point moves on the phase plane ($\mathcal{S}$, $\mathcal{T}$) with fixed p, because we can infer the magnitude and direction of the components $\mathcal{S}_t$ and $\mathcal{T}_t$ of the phase velocity by how close the phase point is to the equilibrium curves and on which side

TABLE 1. Critical p_c as function of δ and C (10^{-5} cm s^{-1}).

	δ		
C	0.1	1.0	5.0
0.025	0.2593	0.2512	0.2431
0.05	0.4537	0.4460	0.4294
0.10	0.7454	0.7373	0.7211
0.20	1.1424	1.126	1.1019
0.50	1.7905	1.766	1.7257
0.75	2.1065	2.082	2.0417
1.00	2.3333	2.309	2.2604

of the curves the phase point lies. The vertical component ($\mathcal{T}_t$) of phase velocity vanishes on the T curve, is negative above it, and positive between the T curve and the origin. The horizontal component ($\mathcal{S}_t$) of the phase velocity is positive between the two branches of the S curve, and negative outside either branch. Knowing the directions of the components enables us to visualize the direction of the phase velocity. As an example, these directions are shown in Fig. 2b. The phase speed is greatest for points far from both curves, and least for points close to both $\mathcal{T}_t = 0$ and $\mathcal{S}_t = 0$ curves.

Note that catastrophe is possible for $p < p_c$. In fact, finite-amplitude perturbations can cause the system to switch between states A and C.

c. *Probabilities of two modes*

The 2×2 model with $\delta = 1$ has a phase hyperspace of eight dimensions, and there are eight hypersurfaces, one for each of the eight equilibrium equations, whose common intersections define the possible equilibrium states. Because of the large number of dimensions in the 2×2 model, a detailed description of the phase speed and direction is not helpful. A Monte Carlo approach, however, can be used to illustrate the relative stability of the two equilibrium states. For a given p, the model is integrated from initial states of random temperatures in each box, between 0° and 25°C and random salinities between 31.5 and 38.5 ppm. (Note that the salinities have to be normalized so that the average salinity for the system remains at 35 ppm for all runs.) The probability of ending up in either of the two stable modes can be computed by numerical integration of the time-dependent equations, as a function of p. Such a figure (for $\delta = 1$) is shown in Fig. 3. For small p the thermal mode is most likely; for $p > 0.44 \times 10^{-5}$ (greater than the critical p_c marking catastrophe) the saline mode is certain. The estimate of present day $p = 0.38 \times 10^{-5}$ cm s^{-1} is marked on the figure by a triangle. It indicates that the thermal mode is more likely; but it lies on a steep slope of the boundary curve and is quite close to catastrophe. Once a system has achieved equilibrium, it may or may not be shifted to the alternate equilibrium mode, depending upon the size of the applied perturbation; however, this is a somewhat different problem than that addressed in Fig. 3.

d. *Transition from a thermal state to a saline state*

An arbitrary point in the phase space defines a state of the system. If at this point the phase speed is zero, then the point represents an equilibrium solution. It is instructive to follow the successive equilibrium states of a system from an initially stable thermal state, through a catastrophe, to the stable salinity-driven state, when the p is increased by a discrete increment. A sample history is shown in Fig. 4. The temperatures T_i and salinities S_i of each box are shown as functions of time in model years. At the bottom, the mass transport u and the phase speed ϕ of the point in phase space ($\phi = \sqrt{\sum_{i=1}^{4} T_{it}^2 + \sum_{i=1}^{4} S_{it}^2}$) are shown.

At time = 0 years the system is in thermally driven equilibrium with precipitation $p = 0.004$ (0.3×10^{-5} cm s^{-1}). The surface equatorial box 1 has a high temperature T_1 and high salinity S_1, and low density σ_1. The surface polar box 2 has a low T_2 and S_2 and a high σ_2. The two deep boxes 3 and 4 have temperatures and salinities equal to those in box 2. The mean temperature over all four boxes is 7.41°C. The circulation u is positive, and the phase speed $\phi = 0$. At time = 40 years the amplitude of the precipitation p is abruptly raised to $p = 0.006$ (0.5×10^{-5} cm s^{-1}). The position of the equilibrium solutions changes abruptly in phase space, but since the phase point itself is still in the old position it moves with a finite phase speed due to the redefinition of the phase space that has occurred with the change

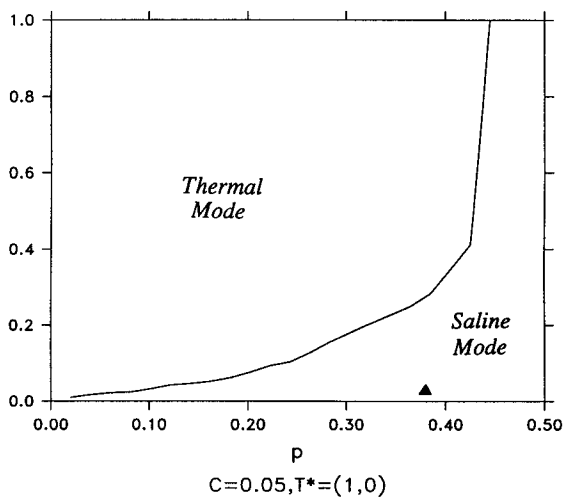

FIG. 3. Probability of reaching the equilibrium modes from random initial conditions in 2×2 box model as a function of amplitude of precipitation p, in 10^{-5} cm s^{-1}. Triangular index marks value of p equivalent to present-day North Atlantic.

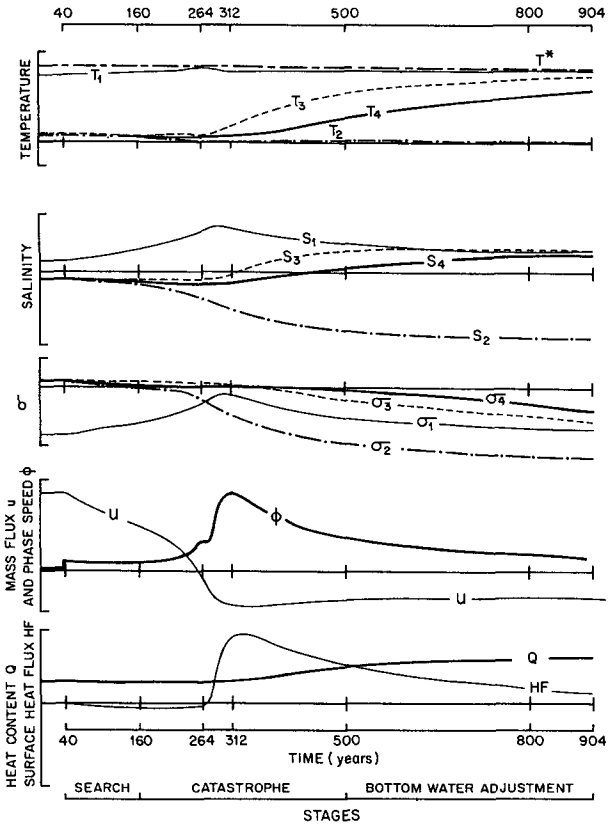

FIG. 4. Evolution of the 2×2 ($\delta = 1$) system from thermally driven equilibrium state to a salinity-driven equilibrium state, as caused by a discrete increase of p at time 40 years, so that p exceeds the critical value. The evolution passes through three stages, as described in the text.

in p. However, it does not increase to a large value because it is still close to the redefined equilibrium hypersurfaces $dT_i/dt = 0$, $dS_i/dt = 0$, although now they do not have a nearby common intersection.

The system responds in three stages: 1) a *searching stage,* 2) a *catastrophic stage,* and 3) an *adjustment stage* during which the deep boxes adjust their temperatures and salinities to the new values that equilibrium requires. For the first 120 years, during the searching stage, the state moves toward higher T_1 and S_1 with decreasing u, seeking the equilibrium anticipated by the increased opposition it encounters from the increased salinity forcing. It moves through a trough of low phase speed ϕ in the phase space because it is close to the equilibrium surfaces. It is looking for the now evanescent common intersection of these surfaces, but cannot find it. One can think of a "strange attractor–repellor" being in this region of hyperspace. When the initial state of the model is very close to the critical state, the search stage can be very long because the phase speed is extremely small when the state is very close to an imaginary equilibrium one. In the case under study, this stage ends at time = 160 years when the phase speed starts to accelerate.

With the search abandoned, the system now enters the catastrophic stage. The phase speed picks up as the point moves out of the deepest part of the trough. The mass flux u drops rapidly through zero, at time 264 years. At this instant the fluxes through the top surface dominate the time changes in T_i and S_i (for $i = 1, 2$). The deep boxes are largely cut off from change. Therefore, it is the time constants of the changes due to surface flux in the top two boxes that determine the duration of the catastrophic stage. In Fig. 4 $\delta = 1$, so we also ran the program with $\delta = 5$ and found that the time required to pass through the catastrophic state was not as strongly affected as for the deep adjustment stage. Therefore, it does seem appropriate to think of the thermohaline catastrophe as a surface phenomenon, geometrically dependent only on the depth of the upper boxes, and not strongly on the thickness or number of deep boxes in the vertical. The phase speed ϕ increases rapidly through a maximum and the mass flux u reverses sign and overshoots too. We are now in the salinity-driven state. The time of ending of this catastrophic stage is rather indefinite, but in the case under consideration occurs approximately at time 500 years, when the mass flux u has reached a close approximation of its final equilibrium value. However, the deep boxes have not yet had time to adjust to the new equilibrium state. At the bottom of Fig. 4 the values of total heat content Q and net heat flux HF from the air to the water are shown as functions of time. The main feature of the heat content Q is the great increase as the system moves from temperature-driven to precipitation-driven equilibrium state, due to heating of the bottom water. There is an increase in salinity of the deep water too, but the total salt remains constant so the mean surface salinity $\frac{1}{2}(S_1 + S_2)$ drops noticeably.

The heat flux HF at first is negative (from ocean to air) as the u decreases during the search stage, but then becomes strongly positive as the deep water absorbs heat from the surface boxes.

The adjustment stage, therefore, begins at time 500 years, and afterward the deep boxes adjust their properties very slowly indeed: box 4 is the slowest. The new equilibrium u, in addition to having opposite sign, is also smaller in absolute value. The phase speed decreases slowly throughout the adjustment stage, and evidently does not reach close to zero (equilibrium) for several thousand years, and even longer if $\delta > 1$.

When equilibrium is finally achieved, the deep box properties are equal to those of box 1 instead of box 2 as before in the thermally driven stage. The mean temperature over all four boxes is 18.15°C (this would be larger with larger δ). Thus, more heat is stored in the ocean than before, and this has had to come through the surface. Accordingly, during the adjustment stage

TABLE 2. Numerical values in the boxes and of maximum flux and phase velocity at critical times.

			$\alpha = 0.16$ $\alpha T^* = 4$				$\beta = 0.771$ $\beta S_0 = 26.99$					
			$-\alpha T_1$	$-\alpha T_2$	$-\alpha T_3$	$-\alpha T_4$	βS_1	βS_2	βS_3	βS_4		
t	Time (years)	Stage	(σ_t units)				(σ_t units)				u	ϕ
0	0											
5	40	search	3.62	0.37	0.37	0.37	27.69	26.75	26.75	26.75	−0.113	0.003
20	160		3.77	0.24	0.35	0.31	28.40	26.32	26.67	26.54	−0.065	0.199
33	264		3.98	0.03	0.32	0.25	29.41	25.60	26.57	26.36	−0.003	0.731
39	312	catastrophe	3.84	0.01	0.85	0.30	29.60	24.89	27.03	26.46	−0.050	1.690
62.5	500		3.87	0.04	2.64	1.30	28.69	23.76	28.22	27.27	−0.039	0.754
100	800	bottom water	3.89	0.07	3.49	2.60	28.29	23.29	28.35	28.00	−0.035	0.354
113	904	adjustment	3.89	0.08	3.62	2.91	28.23	23.28	28.31	28.11	−0.035	0.276

the average surface temperature is lower than in either equilibrium—a slightly paradoxical consequence of the overall oceanic heating. Because the salinity of the deep boxes is greatly increased, and the total salt content remains the same as before, the new equilibrium state has a significantly lower average surface salinity. The meridional gradient of surface temperature is slightly smaller than before, and the meridional gradient of surface salinity is much larger. Table 2 gives the numerical values of the properties in the boxes and of the mass flux and phase velocity at the critical times.

The salinity-driven equilibrium state that the system is now in is very stable to small perturbations. Reduction of p in small steps does not restore the thermally driven state until $p < 0$ at which instant a backward catastrophe occurs toward the thermal state.

A concise way of viewing these dependencies of state upon p (in the 2×2 box with $\delta = 1$) is shown in Fig. 5. The two solid curves depict equilibrium states as functions of p. It will be noticed that these curves have terminations at which one-way transitions to the other state occur. The exact location depends upon the dp, and we have depicted these transitions for small dp. The arrowheads indicate the directions along which the equilibrium state of a system will move as p is raised and lowered.

There is a narrow wedge in the $u^+ - p$ plane, $-p/2 < u^+ < 0$, which corresponds to region III in Fig. 2. As discussed above, within this region water in both layers moves (slowly) southward and carries salt equatorward. Therefore, no steady solution is possible in this region. The existence of this forbidden region is the fundamental reason for the thermohaline catastrophe.

In a typical thermohaline system the haline forcing is against the thermal forcing. When the haline forcing is weak, the thermal forcing dominates and the surface layer moves poleward. As the haline forcing increases gradually, it works as a brake and slows down the poleward motion in the upper layer. Clearly, if the haline forcing is very strong, the circulation should be in the opposite direction. However, there is the forbidden region around $u^+ = 0$, where no steady solution is permissible. Therefore, as the freshwater flux is gradually

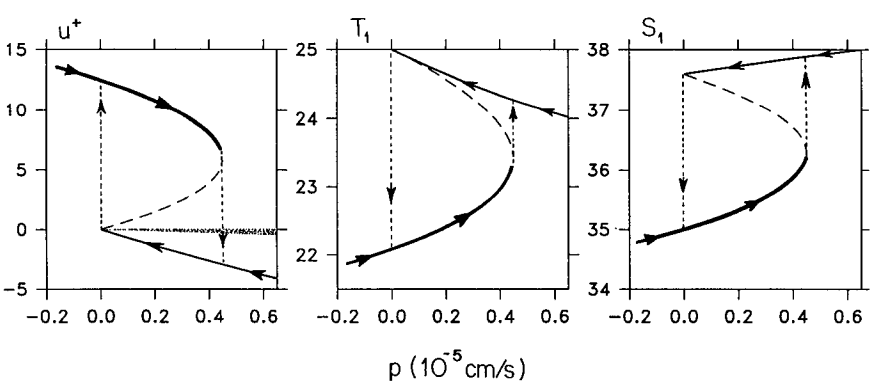

FIG. 5. Dependence of the equilibrium values of u^+ (in units of 10^{-5} cm s^{-1}), T_1 and S_1 upon p, for the 2×2 ($\delta = 1$) box. Heavy solid curve for p increasing by small increments, thin solid curve for p decreasing (thus illustrating the hysteresis of the system), long-dashed lines for an unstable state. The short-dashed lines indicate abrupt transitions.

increased, the steady state of the system must go through a catastrophe in order to reach the opposite circulation pattern. Similarly, if the freshwater flux is reduced gradually, one expects that the system will eventually fall back to the thermal mode. However, due to the existence of the forbidden region, the transition from the haline mode to the thermal mode has to be catastrophic.

The essential ingredient for thermohaline catastrophe is the difference in upper boundary conditions for the temperature and salinity; that is, a large Rayleigh coefficient or a flux condition of the salinity. As discussed by Stommel (1961), if a Rayleigh condition is applied to the salinity with a relaxation coefficient equal or close to the one for the temperature, there would be no multiple solutions nor thermohaline catastrophe, even though the system is still nonlinear.

Although the argument above asserts the necessity of thermohaline catastrophe, it does not tell exactly where the system must collapse. In order to find the critical value p_c, one has to examine the heat and salt balance. Since the transport is a linear of the temperature and salinity difference between two boxes, Eq. (8) is quadratic in $\mathcal{T}$ and $\mathcal{S}$. By eliminating $\mathcal{T}$, a single cubic equation can be obtained for $\mathcal{S}$

$$\mathcal{S}^3 + [2p\mathcal{S}_0 - (1+p)\mathcal{T}_0]\mathcal{S}^2 \pm \frac{p}{c}\mathcal{S}_0\mathcal{S} \pm \frac{2p^2}{c}\mathcal{S}_0^2 = 0 \quad (13)$$

where the upper sign is for region I (thermal mode) and the lower sign for region II (haline mode). The corresponding temperature and velocity can be calculated from (11), (12), and (9).

Equation (13) is cubic for $\mathcal{S}$. For $p < 0$, there are three real roots in region I. However, only one of these roots corresponds to a physical solution, and the other two roots, corresponding to a negative u^+, violate the basic assumption for this region. For $p > 0$, there is one real root corresponding to a negative u^+, so it is nonphysical. For $0 < p < p_c$, the other two roots are real and they correspond to physical solutions. However, only one of them (indicated by heavy line in Fig. 5) is stable and the other one (indicated by long-dashed line) is unstable. At $p = p_c$ these two roots coalesce. For $p > p_c$, they become a pair of complex conjugate roots, and the other real root in region I is a nonphysical solution, as discussed above.

Within region II, (13) has only one real root for any given p. For $p < 0$, the real root corresponds to a positive u^+, so it is nonphysical. Thus, there is just one haline mode for $p > 0$.

The hysterical behavior of the system can be understood better after knowing the algebraic property of the model as discussed above. As p is gradually increased from $p \leq 0$ (with a fairly small dp), the meridional circulation slows down, but the system remains in the thermal mode. As p is increased beyond p_c, the system must fall onto the haline mode. If the system is in the haline mode, it will remain there as long as p is gradually changed. When p becomes zero or negative, the system must fall to the thermal mode.

For $0 < p < p_c$, the system can be forced to jump between the stable thermal mode and the haline mode by a finite amplitude of salinity perturbation. Since the study by Bryan (1986), the catastrophic transition from the thermal mode to the haline mode (the so-called halocline catastrophe) has received much attention. According to our model, given a finite-amplitude salinity perturbation, a catastrophic transition from the saline mode to the thermal mode is also possible.

During the entire sequence of events with an increase of p shown in Fig. 4, there is no gravitational instability in vertical columns of the boxes except, in a weak way, late in the adjustment stage: after 3200 years. On the other hand, the reverse catastrophe, from salinity-driven to the thermally driven state when p is reduced below $p = 0$, is accompanied by extensive gravitational instability in the column with boxes 1 and 3, as the excessive salt in box 1 is dumped down into box 3 by gravitational instability. However, this does not seem to be the place to pursue these events.

The 2×2 model is so highly truncated that when circulating, it cannot have more than one type of water in the two bottom boxes, nor can the forcing represent anything but the latitudinally grand modes. Recent work on the mapping of net freshwater flux (precipitation − evaporation) at the surface of the North Atlantic (Schmitt et al. 1989) shows that the freshwater flux has a markedly different form of latitudinal distribution than the thermal forcing has. Therefore, it seems advisable to try to extend the box model to a 3×2 one (3 horizontal, 2 deep), so that this difference in the functional form of the freshwater forcing can be represented, albeit crudely.

3. The 3×2 box model

The boxes in the upper layer are denoted as boxes 1, 2, 3 (from south to north), and those in the lower layer as boxes 4, 5, 6. Adding these two boxes introduces different flow patterns; we call them modes of circulation. Since there are only two layers in the present case, each mode of circulation is uniquely defined by the locations of sinking. The temperature forcing will be denoted by $[T_1^*, T_2^*, T_3^*]$ and the precipitation forcing by $[p_1, p_2, p_3]$ in which $p_1 + p_2 + p_3 = 0$ and $p = \Sigma_3 |p_i|/2$.

We need some simple way of depicting these modes of circulation and have found it convenient to indicate the mode by the following notation. If we write the normalized magnitude of the three vertical velocities at the bottom of boxes 1, 2, 3 in sequence as, for example, $(1, \epsilon, -1)$, we mean that upwelling is strong

under box 1 (at the equator), weak at box 2 (midlatitude), and that all the sinking occurs at box 3 (the polar box). This is the expected form of the thermal mode of circulation, for $T_i^* = [25, 12.5, 0]$. A saline mode of circulation, with sinking at the equator and smaller amplitude might be written as $(-2\epsilon, \epsilon, \epsilon)$.

Because, in general, the functional form of both thermal and freshwater forcing can be arbitrarily specified, there is a considerable complexity of response. For economy of presentation we limit our discussion to three cases.

Case 1: Linear thermal forcing T = (25, 12.5, 0) and freshwater forcing (−p, 0, p).* Figure 6 is the 3 × 2 counterpart of the 2 × 2 whose probability diagram and mode path map have already been shown in Figs. 3 and 5. Comparing the figures shows that for grand-mode forcing the introduction of the extra pair of boxes (i.e., higher resolution) does not essentially change the response, although more modes of circulation are introduced by the increased number of boxes.

The top panel of Fig. 6 is what we call a mode path map. The abscissa is p. The ordinate is marked with symbols indicating the mode of circulation. For example, the top label is $(1, \epsilon, -1)$, indicating the order of magnitude of the three vertical velocities that define the circulation pattern are upward under box 1, weakly upward under box 2, and downward under box 3. It will be noticed that when $p \to 0$ (purely thermal driving) there are four modes of circulation, whereas in the limit of large p (predominantly saline driving) there is only one. At $p = 0^+$ only the $(1, \epsilon, -1)$ model has a finite circulation throughout all the boxes of the model. In the $(-\epsilon, 1, -1)$ model (with $p \to 0$) the water in the column 1, 4 is stagnant; in the $(1, -1, \epsilon)$ mode column 3, 6 is stagnant, and the whole model is stagnant (with $p \to 0$) in mode $(-2\epsilon, \epsilon, \epsilon)$. These modes in which part, or all, of the model is stagnant when $p \to 0$ are artifacts of the high truncation of the model. They correspond to states in which surface boxes have identical temperatures to that of the surface forcing, $T_i = T_i^*$ ($i = 1$ to 2), and the temperatures and salinities of the boxes beneath them *happen* to be just exactly what is required to reduce the horizontal pressure gradient to zero. The probability of such pressure balances springing up spontaneously from random initial values is reduced as the number of boxes is increased in the vertical. If the initial conditions are broadened to include densities much greater than those formed by the surface forcing, then systems with many superposed layers of boxes can settle into stagnant fossil states where the deep layers are isolated from steady surface forcing. Small random variability in the surface forcings will eventually erode all these stagnant thermal modes and the system will return to the $(1, \epsilon, -1)$ thermal mode of circulation. With $p = 0$ and steady thermal forcing (plus an infinitesimal random fluctuation),

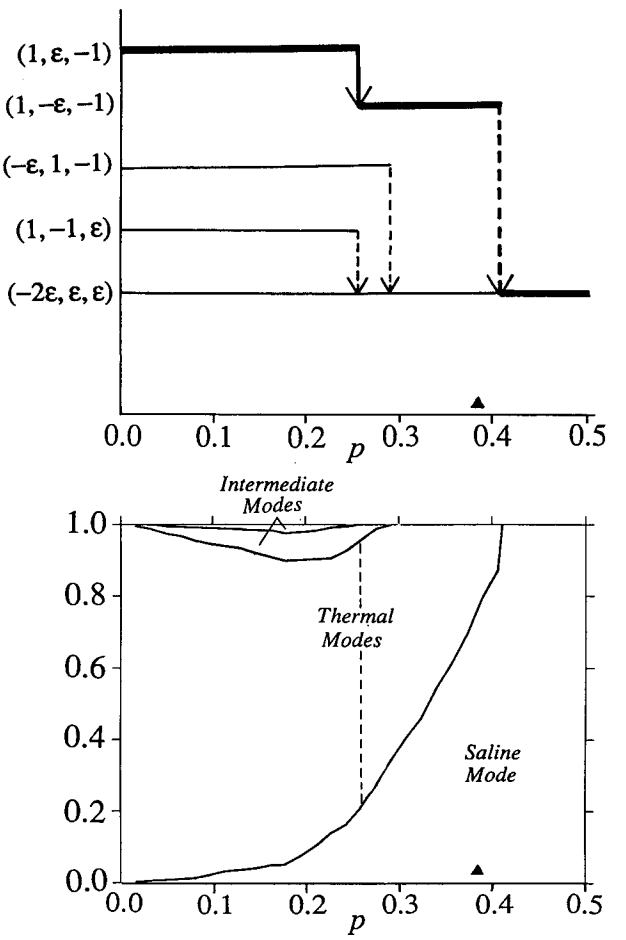

FIG. 6. *Case 1:* 3 × 2 counterpart to 2 × 2 linear freshwater forcing. Path maps (upper panel) and probability of reaching of equilibrium modes from randomized initial conditions of the 3 × 2 box model (lower panel). The ordinate of the path map has discrete levels indicating mode of circulation. Vertical arrows indicate transitions from one mode of circulation to another as p increases. Solid arrows indicate smooth transitions; dashed arrows indicate abrupt transitions. The thickness of the pathlines is meant to indicate degree of probability. In the lower panel there are five possible modes, one smooth transition (thermal modes), and three catastrophes.

The triangular index shows present-day amplitude of p, in 10^{-5} cm s^{-1}. As p is increased from $p = 0$, the main thermal mode eventually collapses to the saline mode $(-2\epsilon, \epsilon, \epsilon)$. If p is now decreased very slowly, the system can be made to remain in the saline mode all the way to $p = 0$.

these stagnant modes are pathological and improbable. On the other hand, it is not difficult to understand why stagnant modes cannot exist at high values of p: there is no way for the surface boxes to reduce the buoyancy flux imposed by inexorable precipitation/evaporation.

Now examining the midportion of the upper panel of Fig. 6, we realize that as p increases slowly the system will move to the right along the heavy most probable path $(1, \epsilon, -1)$. Although for small p the most favored

mode of circulation is $(1, \epsilon, -1)$, this switches smoothly to $(1, -\epsilon, -1)$ as p is increased, through $p = 0.25$, indicating that the opposing freshwater flux succeeds in forcing midlatitude sinking and producing an intermediate water mass, so that the properties in the bottom row of boxes are not uniform. This marks a bifurcation in a graph of bottom temperature versus p (not shown), and we can speak of intermediate water formation. Following this new mode as p increases, at $p = 0.41 \times 10^{-5}$ cm s^{-1} there is a catastrophe (identifiable with that in Fig. 5) and a jump, marked by the one-way dashed vertical line to the saline mode $(-2\epsilon, \epsilon, \epsilon)$.

The lower panel of Fig. 6 is to be compared to Fig. 3. The probability distribution figure is generated by a Monte Carlo method described above. Taking the thermal modes as a whole, the relative probability of falling from random initial conditions into one of the thermal modes, as against the saline mode, is not very different from the probabilities for the 2×2 case in Fig. 3, although the saline mode is somewhat more favored. The modes of circulation $(-\epsilon, 1, -1)$ and $(1, -1, \epsilon)$ are rather improbable: these are labeled intermediate modes in the figure; however, the $(1, -1, \epsilon)$ mode does not produce intermediate water in the oceanographic sense because there is only one column with sinking. The estimated present-day value of p, indicated by the triangle in the figure, shows that $(1, -\epsilon, -1)$ is the expected thermal mode of circulation for the North Atlantic, but that the saline mode is more probable. This inversion of probabilities is the main difference between Figs. 3 and 6 and is due to the effect of truncation on the representation of the grand-mode forcing that is applied. However, one should remember that case 1 does not represent the form of applied freshwater flux realistically, according to Schmitt et al. (1989). The conclusion that we derive from the comparison of the 2×2 and 3×2 model with similar surface grand-mode form of forcing is that finer resolution does not change the character of the solution qualitatively except to introduce improbable minor modes of circulation.

Case 2: Nonlinear freshwater forcing $(-\epsilon p, -p, p)$. At this point we introduce a markedly different form of freshwater forcing, to conform with the recent mapping of evaporation–precipitation over the Atlantic (Schmitt et al. 1989). Instead of choosing a form of $(-p, 0, p)$ as in case 1, we choose $(-0.18p, -0.82p, p)$. This form means that the maximum evaporation is over the midlatitude box instead of the equatorial one. The temperature forcing remains the same as in case 1.

The path map and probability of the different modes of circulation are shown in the two panels of Fig. 7, the present-day value of p marked by the triangle. The situation is quite different from that in case 1 (notice the expansion of the scale of p in the diagrams). The

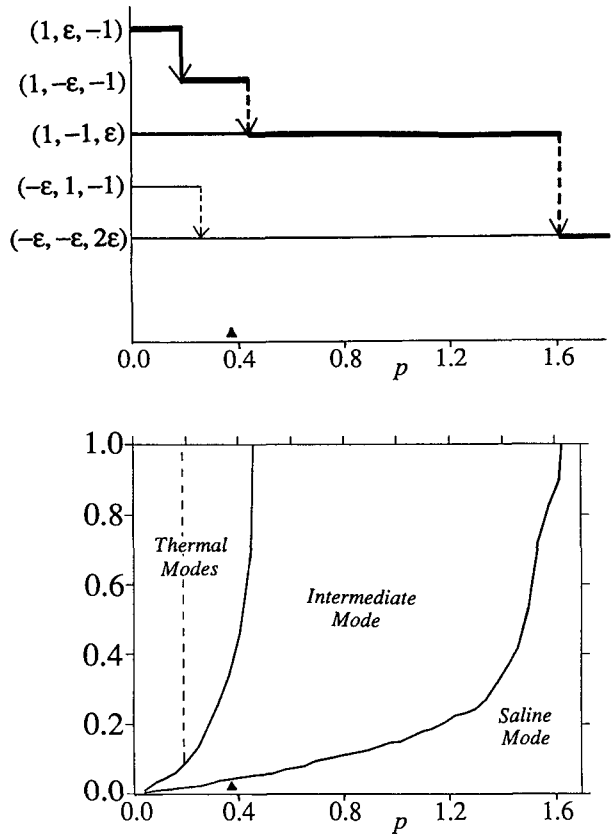

FIG. 7. *Case 2:* North Atlantic freshwater forcing. Path map and probability of mode for case with distribution of p_i chosen to represent freshwater forcing displaced northward (Schmitt et al. 1989). Note change of scale of abscissa. Triangular index shows present-day amplitude of p, which lies in nearly symmetric thermal mode with small-amplitude midlatitude sinking $(1, -\epsilon, -1)$, and close to catastrophe to a strongly asymmetric mode $(1, -1, \epsilon)$, with two meridional cells: thermal in south, saline in north. However, the system is safely remote from the catastrophe to the main saline mode $(-\epsilon, -\epsilon, 2\epsilon)$.

system is most likely to be in the thermal mode of circulation $(1, -\epsilon, -1)$, so that there are two types of deep water (in oceanographic parlance, the less dense of these might be called intermediate water). The present-day value of p lies close to a catastrophe; however, this is not an abrupt transition to the saline mode, but to the intermediate mode of circulation $(1, -1, \epsilon)$, with a strong thermal cell at low latitude and a weak saline cell in high latitude. This flow pattern yields only one type of bottom water, but it is all formed at midlatitude. In trying to model the ice-age oscillation of the North Atlantic, we had hoped to find a catastrophe between a mode $(1, -\epsilon, -1)$ and $(1, -1, -\epsilon)$ close to present-day amplitude of p. This would be consistent with Boyle's (1990) view that the polar deep-water formation is not entirely cut off during ice maximum. However, we found a catastrophe between $(1, -\epsilon, -1)$ and $(1, -1, \epsilon)$. The pure saline mode $(-2\epsilon, \epsilon, \epsilon)$ is im-

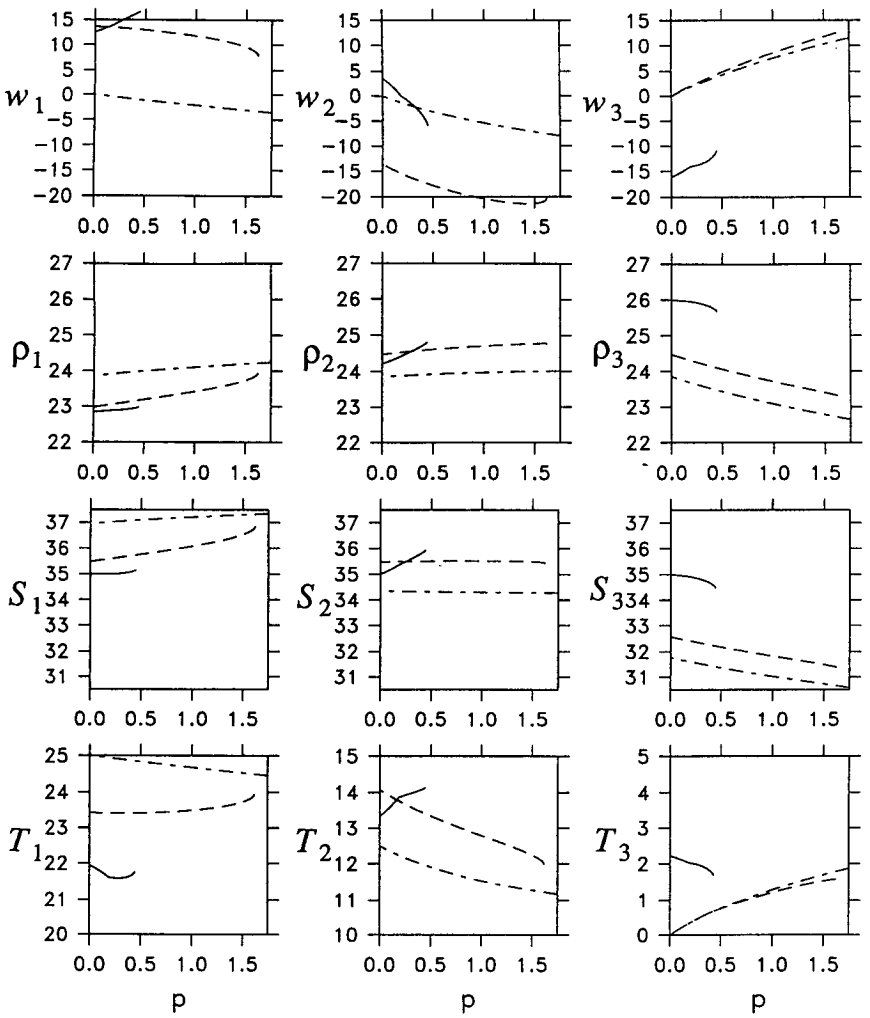

FIG. 8. *Case 2:* North Atlantic freshwater forcing. For $i = 1, 2, 3$ the panels show w_i, ρ_i, S_i, T_i as functions of p. The solid curve shows both thermal modes. The dashed curve shows intermediate mode (with two meridional cells). The dash-dotted curve is the total saline mode.

probable. The catastrophe to the saline mode occurs at unrealistically high amplitudes of p.

Figure 8 displays the vertical velocity w at the bottom of the surface boxes, as well as their density ρ_i, salinity S_i, and temperature T_i for each mode of circulation. When $p = 0$ the system is in the $(1, \epsilon, -1)$ mode with sinking only at the pole, but as the amplitude p is increased, the strong evaporation over box 2 increases the salinity S_2 there and reduces the already small w_2 there so that it passes through zero smoothly and a saline intermediate water is formed at midlatitude (S_5 is not shown in the figure). The system is still in a thermal mode, $(1, -\epsilon, -1)$, and this corresponds to a choice of p representative of the present climate. As p is increased further, the catastrophe to the $(1, -1, \epsilon)$ intermediate mode involves a threefold change in amplitude of w_2 and w_3 (with a change of sign in the latter), larger change in T_1, S_3, T_3, but less in S_1, S_2, T_2. The main change in density is in box 3 where ρ_3 drops 1.6 sigma units.

The dramatic changes in poleward mass flux and the heat flux associated with different modes of circulation are depicted in Fig. 9.

The final catastrophe near $p = 1.6$ suppresses the thermal cell in the southern two boxes. The jumps in ρ_i, S_i, T_i are not as large in this transition, but there are large jumps in w_1 and w_2 with a sign reversal in w_1.

One may interpret these two catastrophes at $p = 0.45$ and 1.6 as a split form of the simple catastrophe at $p = 0.42$ in case 1 (Fig. 6). In case 1 the freshwater forcing has the same form as the thermal forcing, so it must overcome the thermal forcing all at one time. In case 2 the freshwater forcing is stronger near the pole, so it can overcome the thermal forcing in the polar cell first,

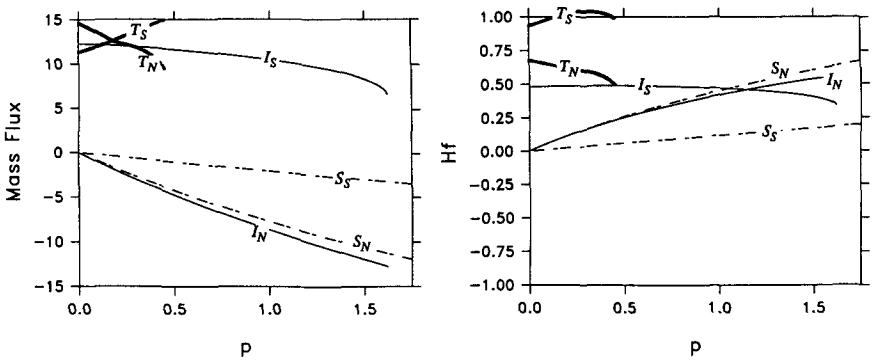

FIG. 9. *Case 2.* North Atlantic freshwater forcing. Meridional mass [in 10^6 m^3 s^{-1} (sv)] and heat (in 10^{15} watts) fluxes as functions of precipitation (in 10^{-5} cm s^{-1}). The notation is: T, thermal; I, intermediate; S, saline; with subscripts N and S to indicate northern or southern cell.

with only a minor effect on the equatorial thermal cell. It is much more difficult for the freshwater forcing (confined mostly to the two polar columns) to reverse the equatorial thermal cell. The second catastrophe is deferred until p is very large. One may speculate that with finer resolution in latitude the catastrophe would be movable in latitude, depending upon p.

Case 3: Nonlinear thermal forcing T^* = [0, 25, 0] and freshwater forcing [p/2, −p, p/2]. To model pole-to-pole modes of circulation, the parameters need to be rescaled to reflect the expansion in the doubling of the horizontal scale. This means setting $c = 0.0222$. The thermal forcing is of form [0, 25, 0] and the freshwater input is $[p/2, -p, p/2]$. The results of running this model are shown in Figs. 10 through 12; they are rather startling. For small p the thermal mode of circulation $(-1, 2, -1)$ with upwelling at the equator (now the central box) and sinking at the poles is preferred. At $p = 0$ the modes $(-1, 1, \epsilon)$ and $(\epsilon, 1, -1)$ are pathological, with a stagnant polar box. However, at realistic value of p one of either pole-to-pole modes is preferred. These are of the form $(-1, 1, \epsilon)$ or $(\epsilon, 1, -1)$. Each is strongly asymmetric to the equator, consisting of a saline mode in one hemisphere, with a thermal mode in the other. The amplitude of circulation of the thermal mode is greater than that of the saline mode. These pole-to-pole modes are isolated from the other modes for p both greater and less than the present realistic value. As p is increased, the system goes catastrophically into a saline mode $(\epsilon, -2\epsilon, \epsilon)$ with antisymmetric cells with joint sinking at the equator.

It is interesting to notice that the asymmetric modes are favored over certain ranges of forcing, although the system is symmetrically forced. In addition, when the thermal mode collapses to a pole–pole mode, the mass flux and poleward heat flux are actually intensified in one hemisphere. In a sense, the system tries to keep up the total amount of deep water formed. If the deep-water formation is cut off in the Northern Hemisphere,

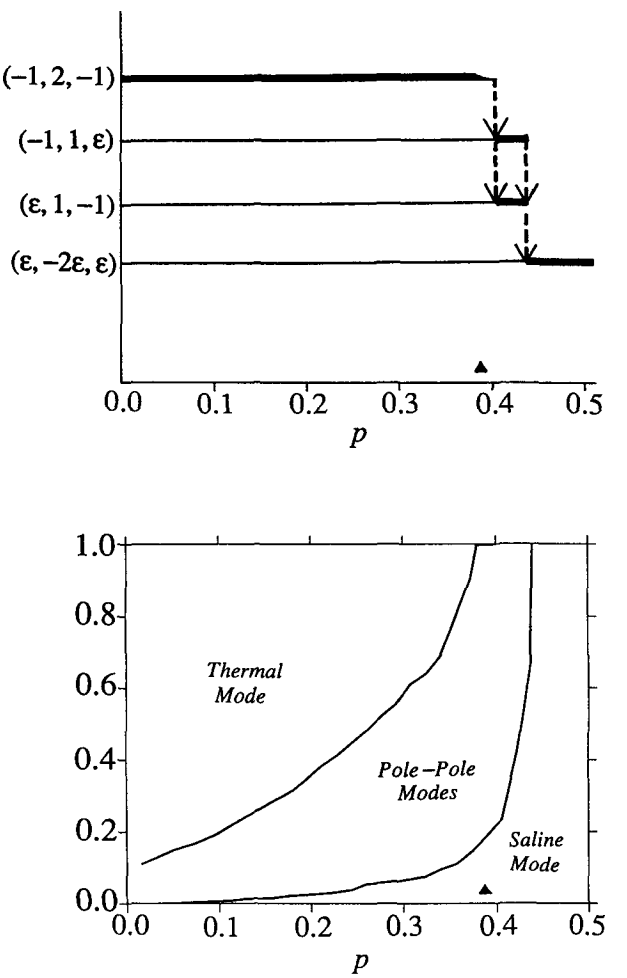

FIG. 10. *Case 3:* Pole-to-pole ocean model with both forcings symmetrical to equator (box 2). The present-day precipitation (triangular index) is close to multiple catastrophes to asymmetrical pole-to-pole modes. They are actually most probable, but neither pole is favored over the other. The pole-to-pole modes have a thermal cell in one hemisphere, a saline one in the other.

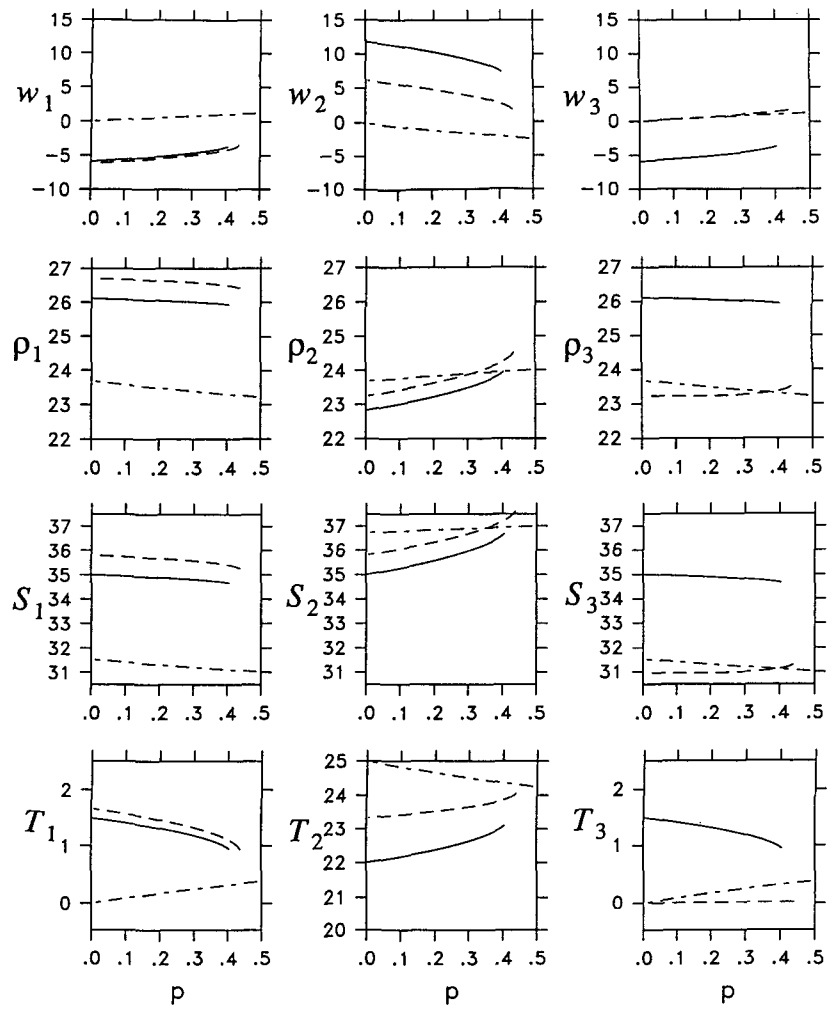

FIG. 11. *Case 3:* Pole-to-pole ocean mode. For $i = 1, 2, 3$ the panels show w_i, ρ_i, S_i, T_i as functions of p. The solid curve denotes thermal mode; the dashed curve denotes one of the intermediate modes; the dot-dashed curve indicates the saline mode.

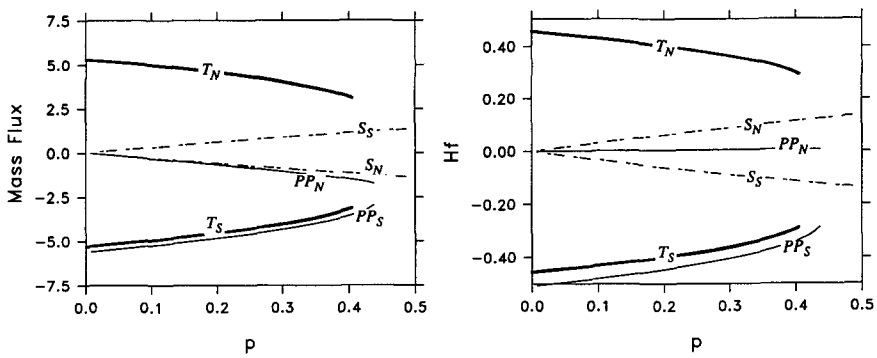

FIG. 12. *Case 3:* Pole-to-pole ocean model. Meridional mass [in 10^6 m^3 s^{-1} (sv)] and heat (in 10^{15} watts) fluxes as functions of p (10^{-5} cm s^{-1}). The notation is: T, thermal; PP, pole to pole; S, saline. The subscripts N and S indicate Northern and Southern hemisphere.

the deep-water formation in the Southern Hemisphere will be slightly increased in compensation. If we could use this model to interpret the paleocirculation during the last glaciation, the slowdown of the deep-water formation in the North Atlantic might induce a slight increase in the amount of bottom water formed near Antarctica and the bottom water flux to the Pacific, Atlantic, and Indian ocean basins, but of course the model is much too crude to claim realism and serves only as a starting point.

4. Conclusions

We have reexamined the structure and the time evolution of the thermohaline catastrophe in a simple 2 × 2 box model. The system has several time scales. For a small perturbation in the forcing, the evolution of the system is determined by the adjustment time scale, which is the inverse of the meridional velocity. However, if the system is very close to the critical state, a small increase of precipitation can cause the system to collapse from a thermal state to a saline state. During a catastrophe, the system responds in three stages: a searching stage, a catastrophic stage, and an adjustment stage. When the initial state of the model is very close to the critical state, the searching stage will be very long because the phase speed of the evolution is small if the state is very close to an imaginary equilibrium one in the phase space.

Using a 2 × 2 model with different lower-layer thickness, it can be seen that the thermohaline catastrophe is a shallow phenomenon, followed by the slower adjustment of the deep circulation.

Although the possible number of multiple equilibrium states increases quickly as the number of boxes is increased, these solutions are not all equally important. Finite-amplitude perturbations can cause catastrophic switching between them. However, for a given set of forcing conditions, there are always one or two solutions that are most likely to appear. Some of the minor mode solutions are very unlikely to appear at all. In general, the solutions of the model can be classified as the grand thermal mode, the grand saline mode, and the minor (or intermediate) modes. Using a value for the precipitation corresponding to the present-day North Atlantic, the modeled thermohaline circulation is very close to a critical state in all four cases. A small increase in the precipitation causes a catastrophic change in the system, from the thermal mode to a saline mode or an intermediate mode.

By imposing on the 3 × 2 model a freshwater flux pattern based on the work of Schmitt et al. (1989), we find that the catastrophe is split in two: the first from the thermal mode to an intermediate mode and the second from the intermediate mode to the saline mode. The present-day thermohaline circulation in the North Atlantic is near the first critical state, and a small increase in the precipitation can cause the thermally dominated circulation of the model to collapse to an intermediate mode, which is characterized by strong sinking at midlatitude and a very sluggish thermohaline circulation in the polar basin. This intermediate mode resembles the circulation during the last glaciation. The critical value required for the second catastrophe (from the intermediate mode to the saline mode) is four times larger than the present-day value. According to the probability distribution study, the saline mode is unlikely to happen.

The Monte Carlo experiment for a pole–pole basin model suggests that the asymmetric modes are the most likely modes for the present-day freshwater flux. However, the symmetric mode with upwelling at the equator is the most likely mode for lower freshwater flux. (Since our model is simplified so much, it is unclear what would happen for a general circulation model.) This observation also reaffirms the meaning of our study on the single-hemisphere models. In fact, the asymmetric mode for a pole–pole basin model can be seen as a superposition of a haline mode and a thermal mode in each hemisphere, as pointed out by Welander (1986). We speculate that the modes, observed in the 3 × 2 single-hemisphere models, may be used as building blocks for a pole–pole basin multiboxed model.

Acknowledgments. R. X. Huang is supported by National Science Foundation Grants OCE88-08076 and OCE90-17158 and Office of Naval Research Grant N00014-90-J-1518. J. Luyten is supported by Office of Naval Research Grants N00014-90-J-1508 and N00014-90-J-1425. H. Stommel is supported by National Science Foundation Grant OCE89-13128.

REFERENCES

Barron, E. J., 1983: A warm equable Cretaceous: The nature of the problem. *Earth-Science Reviews,* **19,** 305–338.

Birchfield, G. E., 1989: A coupled ocean–atmosphere climate model: Temperature versus salinity effects on the thermohaline circulation. *Climate Dynamics,* **4,** 57–71.

——, H. Wang, and M. Wyant, 1990: A bimodal climate response controlled by water vapor transport in a coupled ocean–atmosphere box model. *Paleoceanography,* **5,** 383–395.

Boyle, E. A., 1990: Quaternary deep-water paleoceanography. *Science,* **249,** 863–870.

Broecker, W. S., D. M. Peteet, and D. Rind, 1985: Does the ocean–atmosphere system have more than one stable mode of operation? *Nature,* **315,** 21–26.

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

Duplessy, J. C., N. J. Shackleton, R. G. Fairbanks, L. Labeyrie, D. Oppo, and N. Kallel, 1988: Deep-water source variations during the last climate cycle and their impact on the global deep-water circulation. *Paleoceanography,* **3,** 343–360.

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

Maier-Reimer, E., and U. Mikolajewicz, 1989: Experiments with an OGCM on the cause of the Younger Dryas. *Oceanography 1988,* A. Ayala-Castanares, W. Wooster, and A. Yanez-Arancibia, Eds., UNAM Press.

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

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

——, 1990: Instabilities and multiple equilibria of the thermohaline circulation. Ph.D. dissertation, Ber. Institut für Meereskunde, Kiel, 126 pp.

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

Oppo, D. W., and R. G. Fairbanks, 1990: Atlantic Ocean thermohaline circulation of the last 150 000 years: Relationship to climate and atmospheric CO_2. *Paleoceanography,* **5,** 277–288.

Raymo, M. E., W. F. Ruddiman, N. J. Shackleton, and D. W. Oppo, 1990: Evolution of Atlantic–Pacific $\delta^{13}O$ gradients over the last 2.5 m.y. *Earth Planet. Sci. Lett.,* **97,** 353–368.

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

Schmitt, R. W., P. S. Bogden, and C. E. Dorman, 1989: Evaporation minus precipitation and density fluxes for the North Atlantic. *J. Phys. Oceanogr.,* **19,** 1208–1221.

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

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., D. Reidel, 163–200.

——, 1989: A new type of double-diffusive instability? *Tellus,* **41A,** 66–72.