# Instability of the Thermohaline Circulation with Respect to Mixed Boundary Conditions: Is It Really a Problem for Realistic Models?†

ELI TZIPERMAN,* J. R. TOGGWEILER,** YIZHAK FELIKS,@ AND KIRK BRYAN**

*Environmental Sciences and Energy Research, The Weizmann Institute of Science, Rehovot, Israel
**Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey
@Department of Mathematics, Israel Institute for Biological Research, Ness-Ziona, Israel

(Manuscript received 22 July 1992, in final form 17 November 1992)

ABSTRACT

A global primitive equations oceanic GCM and a simple four-box model of the meridional circulation are used to examine and analyze the instability of the thermohaline circulation in an ocean model with realistic geometry and forcing conditions under mixed boundary conditions. The purpose is to determine whether this instability should occur in such realistic GCMs.

It is found that the realistic GCM solution is near the stability transition point with respect to mixed boundary conditions. This proximity to the transition point allows the model to make a transition between the unstable and stable regimes induced by a relatively minor change in the surface freshwater flux and in the interior solution. Such a change in the surface flux may be induced, for example, by changing the salinity restoring time used to obtain the steady model solution under restoring conditions. Thus, the steady solution of the global GCM under restoring conditions may be either stable or unstable upon transition to mixed boundary conditions, depending on the magnitude of the salinity restoring time used to obtain this steady solution. The mechanism by which the salinity restoring time affects the model stability is further confirmed by carefully analyzing the stability regimes of a simple four-box model. The proximity of the realistic ocean model solution to the stability transition point is used to deduce that the real ocean may also be near the stability transition point with respect to the strength of the freshwater forcing.

Finally, it is argued that the use of too short restoring times in realistic models is inconsistent with the level of errors in the data and in the model dynamics, and that this inconsistency is a possible reason for the existence of the thermohaline instability in GCMs of realistic geometry and forcing. A consistency criterion for the magnitude of the restoring times in realistic models is formulated, that should result in steady states that are also stable under mixed boundary conditions. The results presented here may be relevant to climate studies that run an ocean model under restoring conditions in order to initialize a coupled ocean–atmosphere model.

## 1. Introduction

Ocean models can be run with two alternative formulations of the upper boundary condition (b.c.) for the heat and salt equations. In the first formulation (restoring b.c.), the model is driven by air–sea fluxes of heat and salt that are calculated from the difference between specified surface temperature and salinity fields and the surface temperature and salinity that are calculated by the model. A restoring coefficient (with units of one over time) is used to translate this difference into the heat and salt fluxes driving the model. The second formulation (flux b.c.) involves specifying the air–sea fluxes independently of the values of the surface temperature and salinity.

Bryan (1986) has demonstrated that the steady-state thermohaline circulation obtained under restoring boundary conditions may be unstable upon transition to mixed boundary conditions (i.e., restoring boundary condition for temperature and flux boundary condition for salinity). This instability, and the accompanying phenomena of multiple steady states under mixed b.c. have been studied extensively using models of varying complexities, from box models (Stommel 1961; Rooth 1982; Walin 1985; Marotzke 1989) through intermediate complexity models (Marotzke et al. 1988; Quon and Ghil 1993; Zhang et al. 1993) to primitive equations oceanic general circulation models (Bryan 1986; Weaver and Sarachik 1991; Marotzke and Willebrand 1991).

The implications of this instability to climate modeling may be serious, as indicated for example by Weaver and Sarachik (1991). Climate studies of the response of the atmosphere–ocean system to the doubling of the atmospheric $CO_2$ concentration, for example, are often started by running an ocean model to steady state under restoring conditions using the ob-

† Contribution No. 65, Department of Environmental Sciences and Energy Research, The Weizmann Institute of Science.

*Corresponding author address:* Dr. Eli Tziperman, The Weizmann Institute of Science, Department of Environmental Sciences, 76100 Rehovot, Israel.

served surface temperature and salinity. Once a steady state is reached, the oceanic model is coupled to an atmospheric general circulation model, and both models are run together so that the ocean model is driven by the air–sea fluxes calculated by the atmospheric model. Because of the weak feedback between the freshwater air–sea flux and the surface salinity, the oceanic boundary conditions in the coupled system are similar to using mixed boundary conditions for an ocean-only model. An instability upon transition to mixed b.c. may result, upon transition to the coupled system, in a climate instability that is not a result of any $CO_2$ change and is obviously not desired. Such an instability obtained using today's conditions seems unphysical because the oceanic circulation is known to be roughly stable for the last few thousand years. However, the instability with respect to mixed b.c. has so far been demonstrated only for highly simplified and idealized geometries. Moreover, both box model studies (Walin 1985) and GCM experiments (e.g., Weaver et al. 1991) have shown that an ocean model may be either stable or unstable with respect to mixed b.c., depending on the strength of the freshwater forcing. The still remaining question is, therefore, whether the instability with respect to mixed boundary conditions occurs in models of realistic geometry that are forced by the observed surface salinity and that are used for climate studies. If this instability can occur, it is important to find out if and how it can be avoided, in order to prevent all the implied difficulties for climate studies.

In this study we examine the instability upon transition from restoring to mixed conditions by using both a realistic global primitive equations oceanic general circulation model (PE OGCM) used for climate studies, and a simple box model. Our purpose is to demonstrate that while the instability upon transition from restoring to mixed boundary conditions can occur in a realistic OGCM under some circumstances, it can be avoided by a careful and consistent formulation of the model boundary conditions. It is found that the general circulation model solution is near the stability transition point with respect to mixed boundary conditions. This proximity to the transition point allows the model to make a transition between the unstable and stable regimes induced by a relatively minor change in the surface freshwater flux and in the interior solution. Such a change in the surface flux may be induced, for example, by changing the salinity restoring time used to obtain the steady model solution under restoring conditions. Thus, the steady solution of the global GCM under restoring conditions may be either stable or unstable upon transition to mixed boundary conditions, depending on the magnitude of the salinity restoring time used to obtain this steady solution. The restoring time used in the restoring boundary conditions is, therefore, a crucial parameter determining the stability of ocean models with respect to mixed boundary conditions. A consistency condition for choosing this restoring coefficient is formulated for models of realistic geometry and surface forcing. Using this consistency criterion in realistic models of today's ocean should prevent the undesired and physically unacceptable instability of the present day thermohaline circulation under mixed boundary conditions.

We first show (section 2) that the steady solution of the global PE model obtained using restoring b.c. may become unstable upon transition to mixed boundary conditions. We then demonstrate that increasing the restoring time used for the salinity stabilizes the steady solution obtained using restoring b.c., so that it remains stable under mixed b.c. Note that the change in the salinity restoring coefficient is not accompanied by a corresponding change to the reference salinity to which model salinity is restored [the Levitus (1982) salinity is used as the reference salinity]. Thus, the change in restoring coefficient causes a change in the implied surface water flux, and therefore the interior solution. It is this change to the surface flux (and to the interior solution) that causes the transition between stable and unstable regimes under mixed boundary conditions.

The instability upon transition from restoring to mixed boundary conditions is then studied using a simple four-box model of the meridional circulation. A detailed linear stability analysis of the four-box model is used to first analyze the regimes and mechanisms of instability of a single hemisphere model (section 3) and then to examine the stabilizing effect of increased restoring time (section 4). Several regimes of stability and instability are identified, and it is shown that increasing the salinity restoring time may cause a transition from an unstable regime to a stable regime.

The stability analysis of the box model also sheds some light on the linear instability mechanisms of the thermohaline circulation, refining the mechanism suggested by Walin (1985), providing mainly two new insights into the process: (i) The relative importance of the thermal versus saline forcing of the thermohaline circulation can be measured by the parameter $\beta\Delta S/\alpha\Delta T$, where $\Delta T$ and $\Delta S$ are the difference in surface temperature and salinity between the polar and equatorial boxes, and $\alpha$ and $\beta$ are the temperature and salinity expansion coefficients. Walin (1985), and later Marotzke (1989), have used simple two-box models to show that the thermohaline circulation is unstable when the above parameter is larger than one-half. We show here that for a slightly more complex box model, instability may arise for smaller values of the salinity forcing as measured by the above parameter. (ii) While the main instability mechanism in the upper ocean is found to be that suggested by Walin (1985), it is shown that a somewhat different instability mechanism acts in the deep water.

Motivated by our finding that too short restoring times used to calculate the steady state result in this steady state being unstable upon transition to mixed

b.c., we propose (section 5) that the use of short restoring times is in fact inconsistent with the errors in the observed surface temperature and salinity fields being specified, as well as errors in the model dynamics. A consistency criterion for choosing the restoring times in models of realistic geometry and forcing is formulated. This consistency condition is based on the simple observation that increased restoring times imply larger differences between the specified and calculated surface properties. The criterion states that the rms difference between the observed and calculated surface fields should be of the order of the combined error in the observed surface properties and in the model dynamics. We show that when this criterion is used to calculate the salinity restoring time, the steady solution obtained under restoring conditions is also stable under mixed boundary conditions.

We conclude in section 6.

## 2. Is a realistic PE OGCM unstable with respect to mixed boundary conditions?

Let us first examine if a model of realistic geometry, driven by the observed surface temperature and salinity, is unstable upon transition from restoring to mixed boundary conditions. Because of the relevance of the issues discussed here to coupled ocean–atmosphere model studies, we have chosen to use a global primitive equation general circulation model similar to one that has been utilized in several previous studies in which it was coupled to an atmospheric model (Stouffer et al. 1989; Manabe et al. 1991). The model used here was described in detail by Toggweiler and Samuels (1993). The main difference between the present model and that used in the coupled model studies is the horizontal mixing used here versus the isopycnal mixing used in the coupled model studies. The model is based on the model described by Bryan (1969) with later modifications by Semtner (1974) and Cox (1984).

The model was first run using restoring boundary conditions for both the temperature and salinity, using restoring coefficients of 1/30 days [see entry (a) in Table 1], and restoring the surface fields to the Levitus (1982) annually averaged climatological data. The heat ($H$) and freshwater ($E - P$) fluxes driving the model are calculated from the observed surface fields ($T^d$, $S^d$) and the model solution at the uppermost level ($T$, $S$) using

$$H = \rho_0 C_p \gamma_T \Delta z_1 (T^d - T)$$
$$E - P = \gamma_S \Delta z_1 (S^d - S)/S_0, \qquad (1)$$

where $\rho_0$ is a constant reference density, $C_p$ is the water heat capacity, $\Delta z_1$ is the thickness of the uppermost model level, $S_0$ is a reference salinity, and ($\gamma_T$, $\gamma_S$) are the restoring coefficients for temperature and salinity.

The steady-state meridional circulation streamfunction for the section of the global model corresponding to the Atlantic Ocean is shown in Fig. 1a. Diagnosing the implied salt flux at steady state (Fig. 1b) and changing to mixed b.c. (flux b.c. for salinity and restoring conditions for temperature), we find that the steady solution becomes unstable, as commonly found in previous studies for models of idealized geometry. This instability leads to the collapse of the thermohaline circulation (Bryan 1986) within about 100 years. We do not wish to elaborate on the description of the model evolution past the initial instability stage, as it is sufficient for our purpose here to note that the model solution is unstable under mixed b.c.

Next, we increase the restoring time for the salinity field from 30 to 120 days [run (b) in Table 1], and repeat the experiment. The steady meridional circulation streamfunction and the implied salt flux at steady state are given in Fig. 2. This time the solution seemed stable upon transition to mixed b.c. The stability of the solution was examined by running the model under mixed boundary conditions for about 1000 years. Note that the instability in the previous case could be observed after a few years run.

To try to understand these findings, we consider in the following section a highly simplified box model of the meridional circulation. Once the stability of the meridional circulation is carefully analyzed for the simpler model, we return to the global primitive equation general circulation model results in sections 4 and 5.

## 3. Instability of a hemispheric box model of the thermohaline meridional circulation

Our purpose in this section is to first analyze the linear stability of a simple box model of the meridional circulation. We then use this stability analysis in section 4 to explain the mechanism by which increasing the restoring time for salinity results in the stabilization of

TABLE 1. Summary of GCM runs used in this study. The restoring coefficients used for temperature and salinity are given by $\gamma_T$ and $\gamma_S$, and are equal to the inverse of the restoring times. The symbols rms ($\Delta SST$) and rms ($\Delta SSS$) refer to the rms difference between model and data surface temperature and salinity—e.g., rms ($\Delta SST$) = rms ($SST_{\text{model}} - SST_{\text{data}}$). The rms values for temperature and salinity are given in degrees Celsius and parts per thousand, respectively.

| Run | $1/\gamma_T$ (days) | $1/\gamma_S$ (days) | rms ($\Delta SST$); rms ($\Delta SSS$) | Meridional circulation (Sv) | Mixed b.c. | Figure |
|---|---|---|---|---|---|---|
| a | 30 | 30 | 0.54; 0.11 | 18.3 | unstable | 1 |
| b | 30 | 120 | 0.51; 0.22 | 22.3 | stable | 2 |

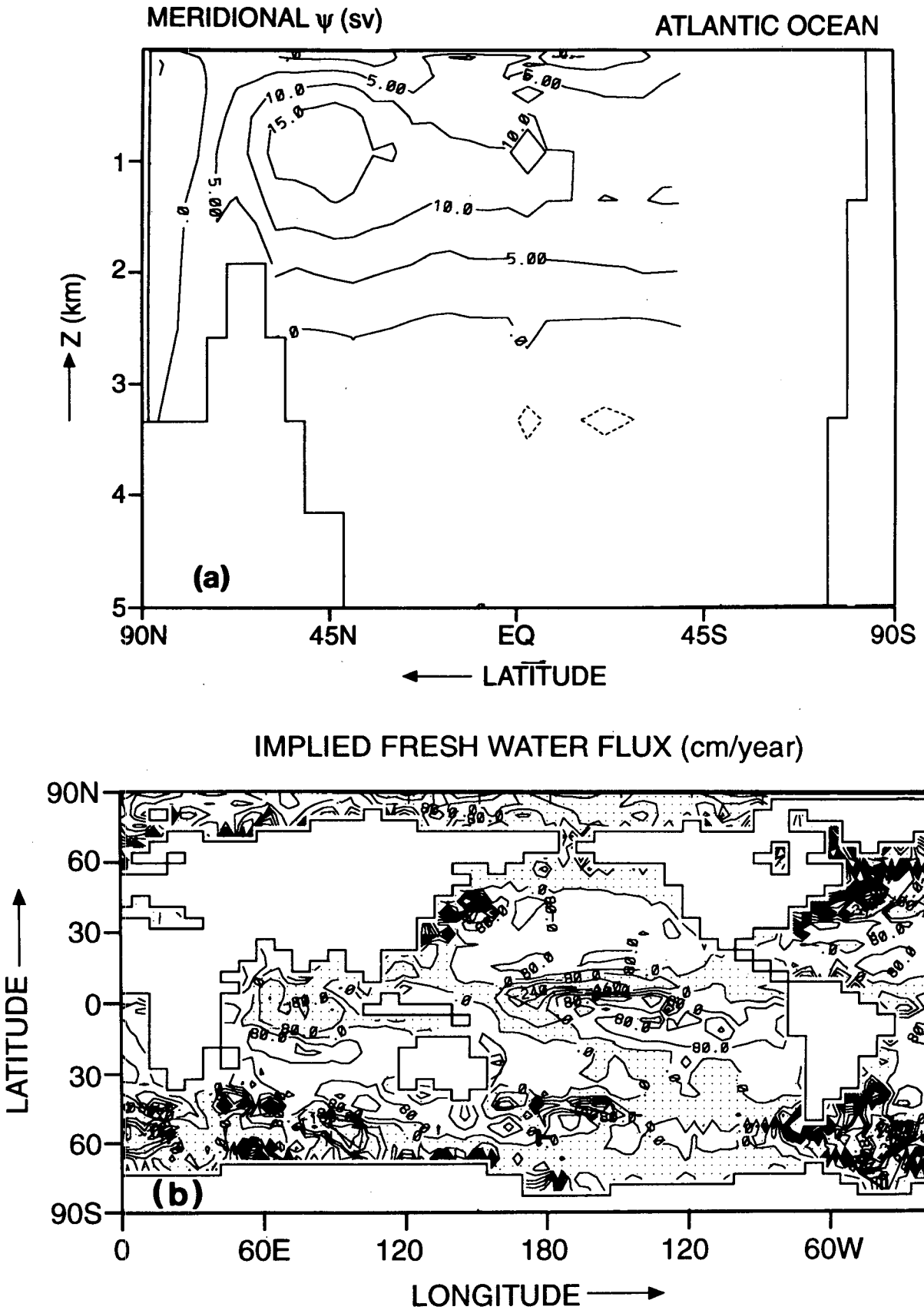
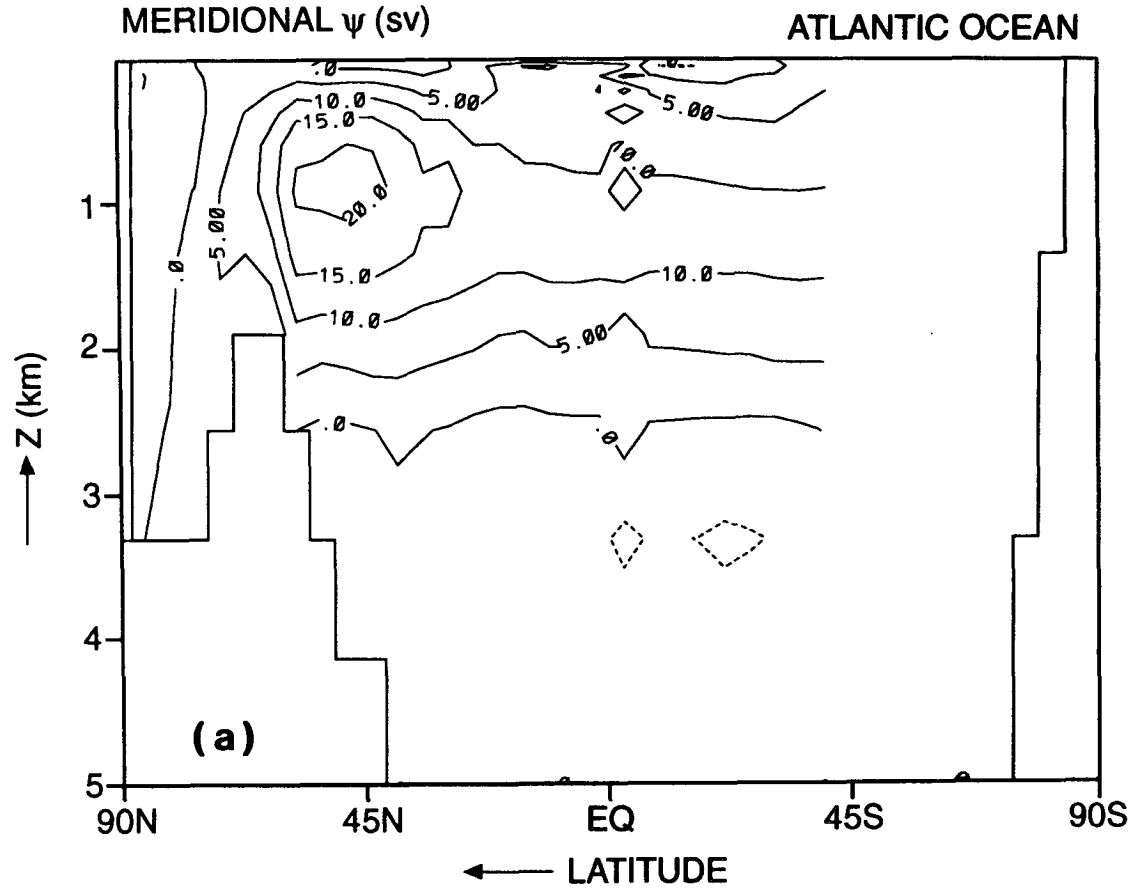



FIG. 1. Solution of the global PE model using restoring times of 30 days for both temperature and salinity. (a) Meridional circulation streamfunction for the Atlantic Ocean (in Sv). (b) Implied surface freshwater flux (cm $yr^{-1}$).



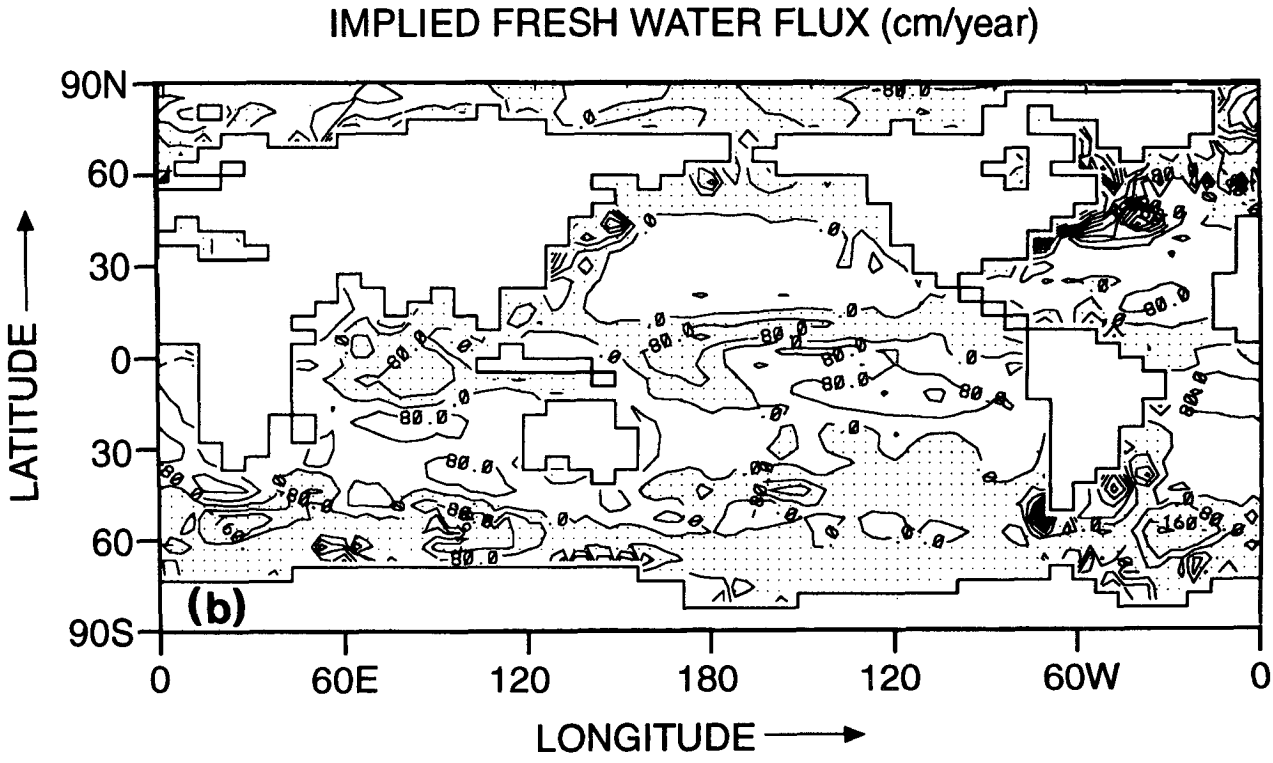




FIG. 2. Solution of the global PE model using a restoring time of 30 days for temperature and 120 days for salinity. (a) Meridional circulation streamfunction for the Atlantic Ocean (in Sv). (b) Implied surface freshwater flux (cm $yr^{-1}$).

the model solution upon transition from restoring to mixed boundary conditions.

The box model is essentially that used by Huang et al. (1992) with two major differences: (i) we allow for different sizes of the polar and equatorial boxes; and (ii) the boundary condition for the salinity involves a salt flux, ignoring the actual mass flux involved, in order to follow the procedure used in GCMs.

The model dynamics include a simple frictional horizontal momentum balance, and are hydrostatic, mass conserving, and with the advection of temperature balanced by a surface flux:

$$0 = -\frac{1}{\rho_0} P_z - \frac{g}{\rho_0}\rho$$

$$0 = -\frac{1}{\rho_0} P_y - rv$$

$$v_y + w_z = 0$$

$$\rho = \rho_0[1. - \alpha(T - T_0) + \beta(S - S_0)]$$

$$T_t + (vT)_y + (wT)_z = 0; \quad \kappa T_z|_{z=0} = \gamma_T(T^* - T). \tag{2}$$

The salinity equation is used in two alternative forms, one with restoring boundary conditions, and the second using flux conditions:

$$S_t + (vS)_y + (wS)_z = 0; \quad \kappa S_z|_{z=0} = \begin{cases} \gamma_S(S^* - S) \\ H_s. \end{cases} \tag{3}$$

In the above equations $(y, z)$ and $(v, w)$ are the (northward, upward) coordinates and velocities; $P$ denotes the pressure; $T$, $S$ the model temperature and salinity; $T^*$, $S^*$ the specified surface temperature and salinity; $H_s$ is the surface salt flux; $g$ is the gravitational acceleration, $\rho_0$ a constant reference density; $r$ is a friction coefficient; $\alpha$ and $\beta$ are expansion coefficients for the temperature and salinity; and $\kappa$ is a vertical mixing coefficient.

These equations are nondimensionalized using the following scales: $L$ and $D$ for the horizontal and vertical distances; $\gamma_T^{-1}$ for time; $R^* = \alpha(T_2^* - T_1^*)$ for density, where $T_2^*$ and $T_1^*$ are the specified temperatures in the polar and equatorial regions, respectively; $DR^*g$ for pressure; and $\gamma_T L$ and $\gamma_T D$ for horizontal and vertical velocities. The thickness ratio between the upper and lower boxes is denoted by $\delta$, while the width ratio between the equatorial and polar boxes is denoted by $\Delta$ (see Fig. 3). Writing the nondimensionalized equations for the four-box model shown in Fig. 3, we obtain the following set of equations (Huang et al. 1992)

$$\rho = -\alpha(T - T_0) + \beta(S - S_0)$$

$$v = [(\rho_4 - \rho_3) + \delta(\rho_2 - \rho_1)]/b$$

$$\delta\frac{\partial T_1}{\partial t} = v\delta(T_3 - T_1) + \delta(T_1^* - T_1)$$

$$\delta\Delta\frac{\partial T_2}{\partial t} = v\delta(T_1 - T_2) + \delta(T_2^* - T_2)$$

$$\frac{\partial T_3}{\partial t} = v\delta(T_4 - T_3)$$

$$\Delta\frac{\partial T_4}{\partial t} = v\delta(T_2 - T_4)$$

$$\delta\frac{\partial S_1}{\partial t} = v\delta(S_3 - S_1) + F_s^1$$

$$\delta\Delta\frac{\partial S_2}{\partial t} = v\delta(S_1 - S_2) + F_s^2$$

$$\frac{\partial S_3}{\partial t} = v\delta(S_4 - S_3)$$

$$\Delta\frac{\partial S_4}{\partial t} = v\delta(S_2 - S_4), \tag{4}$$

where

$$F_s^1 = \begin{cases} \gamma\delta(S_1^* - S_1) \\ H_s^1 \end{cases}, \quad F_s^2 = \begin{cases} \gamma\delta(S_2^* - S_2) \\ H_s^2 \end{cases}$$

$$b = \frac{1}{4}(1 + \delta)(1 + \Delta)\frac{\rho_0 r\gamma_T L^2}{gR^*D}, \quad \gamma = \frac{\gamma_S}{\gamma_T}. \tag{5}$$

The velocity $v$ is the water velocity between boxes 1 and 2. Note that we have used upwind differencing assuming a positive (northward) meridional velocity. This amounts to examining the temperature-dominated circulation, driven by the cooling and sinking at the northern box (upper-right box in Fig. 3), and heating in the south. For negative velocity (salinity dom-

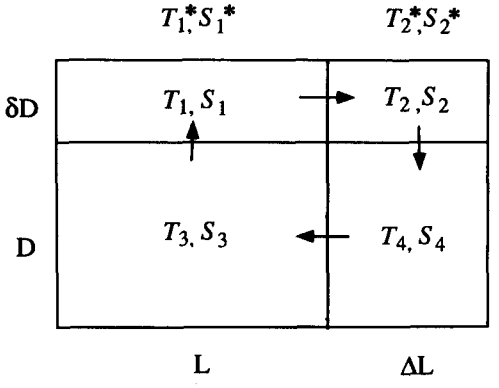



FIG. 3. Schematic plot of the box model geometry.

inated circulation), the finite-difference form changes, although this is not a concern in this work, because we are interested in the stability of today's thermohaline circulation (e.g., in the North Atlantic Ocean), which is temperature dominated.

At steady state, the above set of equations may be used to derive a simple third-order polynomial equation for the difference $T_2 - T_1$. This equation can be easily solved [using routine C02AGF from Numerical Algorithm Group (1984)], and its solution can be used to calculate the temperature and salinity at each box. Of the three solutions of the polynomial equation, only one turns out to be physical under restoring conditions, and two are physical under mixed boundary conditions. [Physical solutions must be real, and with positive northward velocity as assumed in the upwind differencing. A fuller discussion of the bifurcation behavior of model solutions as function of various model parameters is not needed here and is given by Thual and McWilliams (1992) and Quon and Ghil (1992). In addition to the two thermally dominated solutions found here there is a salinity-dominated solution under flux conditions (or under restoring conditions with a long relaxation time for the salinity), as is well known from previous studies (e.g., Stommel 1961; Marotzke 1989).] Thus, we can solve for all steady-state model solutions without using time integration, therefore allowing the calculation of both stable and unstable steady solutions.

### *a. Stability regimes*

Before considering the effect of changing the salinity restoring time, let us consider the linear stability analysis of the model. Let $\bar{T}_i$, $\bar{S}_i$, $i = 1, \cdots, 4$ be a particular steady solution. Considering an infinitesimal perturbation $T'_i$, $S'_i$ and linearizing the model equations (4) about the steady solution using mixed boundary conditions, we obtain a linear set of equations for the perturbation, which can be written in matrix form

$$\frac{\partial}{\partial t}\begin{pmatrix} T' \\ S' \end{pmatrix} = \mathbf{A}\begin{pmatrix} T' \\ S' \end{pmatrix}. \tag{6}$$

The stability of the model is determined by the eigenvalues of the matrix **A**: let

$$\begin{pmatrix} T' \\ S' \end{pmatrix} = \mathbf{X}e^{\lambda t} \tag{7}$$

and substitute this in the perturbation equation, to obtain the eigenvalue problem

$$\mathbf{AX} = \lambda\mathbf{X}, \tag{8}$$

where **X** is an eight vector containing the perturbation temperature and salinity, and **A** is an eight by eight matrix. Clearly, the steady solution is unstable if there exists an eigenvalue $\lambda$ with positive real part [see Marotzke (1990) for a similar analysis of the stability of a box model, with salinity as the perturbed variable]. Similarly, a complex eigenvalue corresponds to an oscillatory behavior of the small perturbation about the steady state.

To identify the different stability regimes of the model, we specify the salinity flux at the surface, solve the steady model equations under mixed boundary conditions to obtain the two physical temperature-dominated steady solutions, and analyze the stability of the solutions by solving the above eigenvalue problem. Following this procedure for different values of the salt flux, we obtain a full characterization of the model stability. There are four types of possible solutions, characterized by the sign and value of the real and imaginary part of the eigenvalues of **A**: (i) stable solutions where small perturbations decay exponentially; (ii) stable solutions where small perturbations decay exponentially and oscillate as well; (iii) unstable solutions where small perturbations oscillate and grow exponentially; and (iv) unstable solutions where small perturbations grow exponentially.

The results of the characterization of stability regimes are given in Figs. 4a–c. The model parameters used to obtain Fig. 4 are $\Delta = 3.0 \times 10^{-2}$, $\delta = 1.0$, $L = 5000$ km, $D = 2$ km, $\alpha = -1668. \times 10^{-7}$ K$^{-1}$, $\beta = 7.61 \times 10^{-4}$ ppt$^{-1}$, $T_2^* - T_1^* = 25$°C, $r = 3. \times 10^{-3}$, $\gamma_T^{-1} = 30$ days, and $g = 980$ cm s$^{-2}$.

It is convenient to examine the solutions as a function of the strength of the salt flux and the surface salinity forcing as measured by the parameter $\mathcal{S} = \beta\Delta S/\alpha\Delta T$ (Walin 1985; Marotzke 1989). (Note that $\Delta S$ is known only once the model solution is found; under restoring conditions using a short relaxation time, $\Delta S$ is nearly equal to the equator-to-pole salinity difference of the specified surface salinity.) In all three Figs. 4a–c, the possible solutions lie on a curve for which each salt flux value corresponds to two possible steady temperature-dominated solutions. In Fig. 4a, the four regimes of instability are indicated by different markers on the curve of possible steady solutions. This figure was obtained by examining both the number of unstable eigenvalues of the matrix **A**, and the number of complex eigenvalues. Figure 4b shows the number of unstable eigenvalues for all possible steady solutions, while Fig. 4c shows the number of complex eigenvalues.

As the salinity forcing, measured by $\mathcal{S} = \beta\Delta S/\alpha\Delta T$, increases, the model undergoes several transitions. First, it is stable to infinitesimal perturbations ("+" in Fig. 4a), then it is oscillatory stable (*), then oscillatory unstable (○), and finally it is unstable (×).

The location of the transition from stable to unstable regimes is a crucial factor in the mechanism we propose below for the stabilizing effects of increased salinity restoring times. Both Walin (1985) and Marotzke (1989) found that the model is unstable for $\mathcal{S} > 0.5$. Note, however, that we obtain a different result here, where the model becomes unstable for smaller values

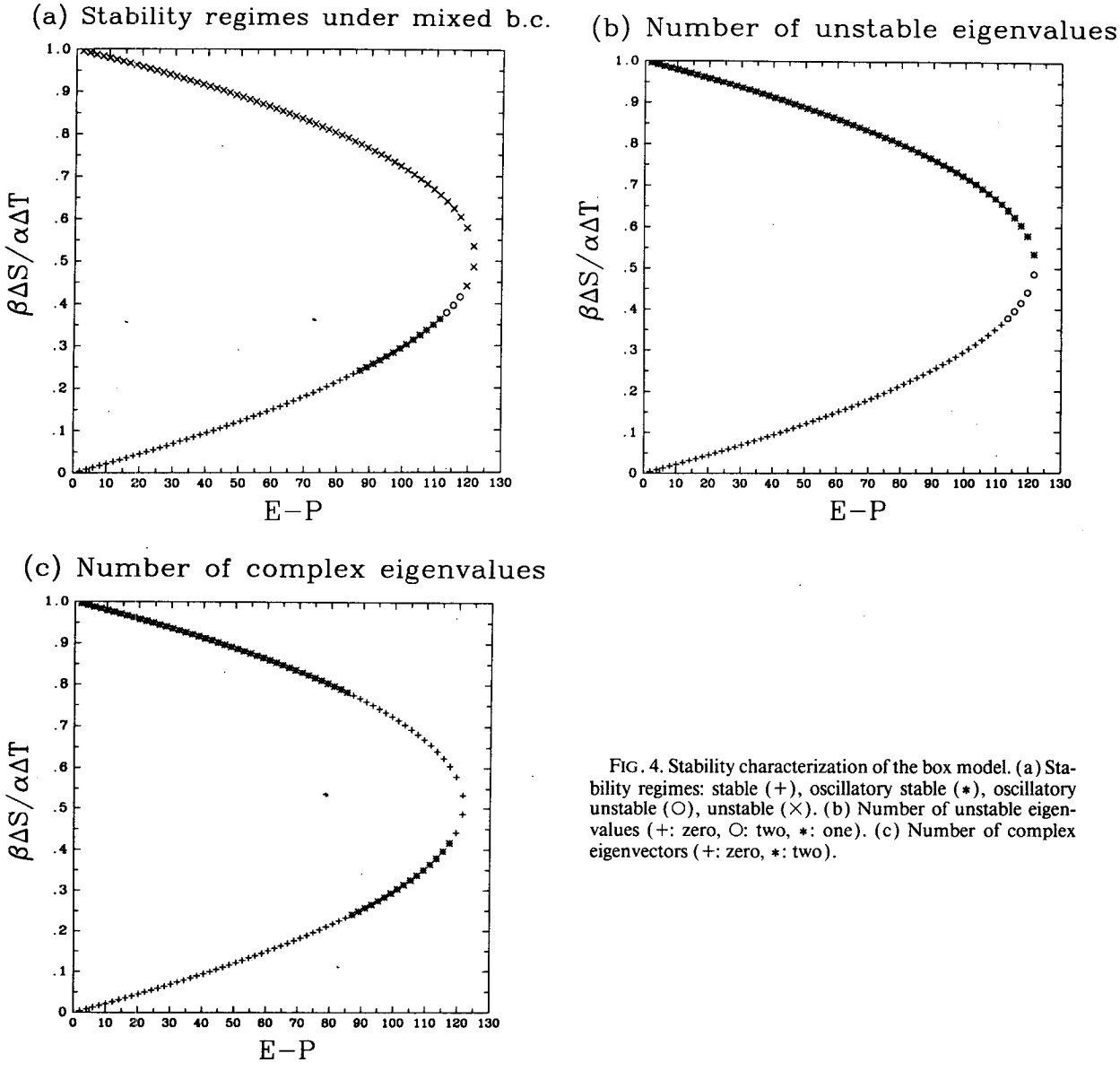



FIG. 4. Stability characterization of the box model. (a) Stability regimes: stable (+), oscillatory stable (∗), oscillatory unstable (○), unstable (×). (b) Number of unstable eigenvalues (+: zero, ○: two, ∗: one). (c) Number of complex eigenvectors (+: zero, ∗: two).

of $\mathcal{S}$. It is in particular important for the later developments that the transition occurs below the maximal freshwater flux for which temperature-dominated solutions exist. This allows for a transition from unstable to stable regimes by a decrease in salinity flux forcing. This decrease in flux forcing also corresponds to a decrease in the salinity forcing measured by $\mathcal{S} = \beta\Delta S/\alpha\Delta T$. Note that in a two-box model such as Stommel's (1961) model analyzed by Marotzke (1989), the stability transition point is at the value of the maximum salt flux for which temperature-dominated solutions exist. This implies that the stability transition in such a model is marked by a decrease in the surface salinity gradient, but not necessarily in the surface salinity flux forcing. We will see below that the transition from unstable to stable regimes in the primitive equation GCM involve a small reduction in the surface salt flux, as in our four-box model.

It seems that the difference in the location and character of the stability transition point between the different box models results from the addition of two boxes relative to the simpler 2-box models used by Walin (1985) and Marotzke (1989). It is possible that adding more boxes will again change the critical value of $\mathcal{S}$. One expects, though, that as the number of boxes increases, the critical $\mathcal{S}$ will converge to some value between 0.5 and zero. This dependence of the critical stability point on the number of boxes is perhaps reminiscent of the bifurcations of box model solutions as demonstrated by Thual and McWilliams (1992). They

have shown that as the number of boxes increases, the catastrophe structure of the box models evolves, and that one needs some minimal number of boxes to obtain the major characteristics found in a continuous model.

### b. *Instability mechanism*

The instability of a symmetric thermohaline circulation under flux boundary conditions has been demonstrated by Rooth (1982), using a simple three-box model. Walin (1985) has described a simple mechanism by which a symmetric thermohaline circulation is destabilized by an antisymmetric salinity perturbation. Marotzke et al. (1988) and Marotzke (1990) further discussed Walin's mechanism, applied it to other than symmetric circulations, and used it to interpret the results of a numerical model.

In this section, the stability analysis of the previous subsection is continued, this time examining the physical mechanisms of linear instability for the regions represented by each of the four boxes of our model. It is shown that while Walin's mechanism acts in the upper part of the ocean, the linear instability of the deep water is governed by a different mechanism.

As mentioned in the previous subsection, the model is unstable to the small perturbation governed by (6) when the matrix **A** has eigenvalues with a positive real part. The eigenvalues of the matrix **A** were used in the previous subsection to classify the stability regimes of the model. But additional useful information may be extracted from the eigenvectors of **A**. These may be interpreted as the temperature and salinity perturbations that grow, oscillate, or decay according to their corresponding eigenvalues. The eigenvector that corresponds to the most unstable eigenvalue, for example, provides the spatial structure of the most unstable temperature and salinity perturbation.

Let us now examine the instability mechanism in each box, by physically interpreting the eigenvector structure. For this purpose we present the different terms in the linearized temperature and salinity equations, as calculated for perturbations that are equal to the elements of an unstable eigenvector. Let us write, for example, the linearized temperature and salinity equations (using mixed boundary conditions) for box 1 as

$$\delta\frac{\partial T_1'}{\partial t} = [v'\delta(\bar{T}_3 - \bar{T}_1)] + [\bar{v}\delta(T_3' - T_1')] - \delta T'$$
$$= [\nabla(v'\bar{T})] + [\nabla(\bar{v}T')] - \hat{\gamma}\delta T'$$
$$\delta\frac{\partial S_1'}{\partial t} = [v'\delta(\bar{S}_3 - \bar{S}_1)] + [\bar{v}\delta(S_3' - S_1')]$$
$$= [\nabla(v'\bar{S})] + [\nabla(\bar{v}S')]. \quad (9)$$

Here $v'$ is the perturbation velocity calculated from the perturbation density, which in turn is calculated from the perturbation temperature and salinity given by the elements of the eigenvector. The factor $\hat{\gamma}$ is equal to one in our nondimensional units, but is kept to make it clear that the source of the term $-\hat{\gamma}T'$ is the restoring boundary conditions for the temperature. The expressions on the second line in each of the above equations represent in a symbolic way the effects of the advection of mean gradients by the perturbation velocity (e.g., $[\nabla(\bar{v}T')]$), and the advection of the perturbation gradients by the mean velocity (e.g., $[\nabla(\bar{v}T')]$). These are merely convenient notations for the accurate expressions in the first line of each of these equations.

Using similar notations for the rest of the boxes, the different terms in the temperature and salinity equations for an unstable eigenvector (with a vanishing imaginary part) can be written, and this allows examining the instability mechanism for both salinity and temperature, and in all boxes.

We begin (Table 2) with the simplest instability regime, where the instability does not involve oscillations, and the unstable eigenvalue is real (regime marked × in Fig. 4a).

The mechanism suggested by Walin (1985), and further discussed by Marotzke et al. (1988), acts as follows. The salinity perturbation creates a perturbation velocity $v'$. This perturbation velocity advects the mean salinity gradients, to enhance the salinity perturbation. The feedback between the growth of the salinity perturbation $S'$ and the perturbation velocity $v'$ creates the exponential growth and eventually destroys the mean flow. This mechanism is clearly based on the destabilization of the flow by the term $\nabla(v'\bar{S})$.

Examining the different terms in the salinity equation given in Table 2 shows that the dominant term for the surface boxes (boxes 1 and 2) is indeed $\nabla(v'\bar{S})$, in agreement with the above mechanism. The perturbation in the deeper boxes (boxes 3 and 4), however, grows using a different mechanism. For these boxes the mean salinity gradients are zero because the salinities of boxes 2, 3, and 4 are equal at steady state. Here the stability mechanism relies on the term $\nabla(\bar{v}S')$, that is, on the advection of the perturbation salinity $S'$ by the mean velocity $\bar{v}$. In the ocean, the properties of the deep water vary very little from the water mass formation sites to the deep water elsewhere, so that it makes sense that the linear instability mechanism for the deep water should indeed not be based on the advection of the mean gradients by the perturbation velocity, but on the opposite mechanism.

Both Walin (1985) and Marotzke (1990) have assumed that the temperature is basically fixed during the instability process, due the fast restoring times for the surface temperature field. While this is a convenient assumption for exploring the basic mechanism of instability, it is obvious that the instability process must also involve an exponential growth of the temperature perturbation. This is because the instability modifies the circulation so that the steady solution for the tem-

TABLE 2. Stability analysis of an unstable eigenvector of solution in the unstable regime with a single unstable mode. This solution is the one on the upper branch of Fig. 4a, derived with $E - P = 111.5$ cm yr$^{-1}$.

| | Box | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| $\bar{T}$ | .99955 | $.14892 \times 10^{-1}$ | $.14892 \times 10^{-1}$ | $.14892 \times 10^{-1}$ |
| $\bar{S}$ | 1.4085 | −.35937 | −.35937 | −.35937 |
| $\bar{v} = .4537 \times 10^{-3}$ | | | | |
| $v'$ calculated from $T'$, $S'$, and $\rho' = -.3673 \times 10^{-4}$ | $.6503 \times 10^{-3}$ | $.6136 \times 10^{-3}$ | | |
| $S'$ | $-.45982 \times 10^{-1}$ | .92854 | $.70940 \times 10^{-2}$ | .36772 |
| $T'$ | $-.59021 \times 10^{-3}$ | $.19389 \times 10^{-1}$ | $.14813 \times 10^{-3}$ | $.76786 \times 10^{-2}$ |
| $\rho'$ | $.16182 \times 10^{-1}$ | .31931 | $-.24395 \times 10^{-2}$ | −.12646 |
| $\nabla(\bar{v}S')$ | $.24082 \times 10^{-4}$ | $-.14739 \times 10^{-1}$ | $.16363 \times 10^{-3}$ | $.84820 \times 10^{-2}$ |
| $\nabla(v'\bar{S})$ | $-.10847 \times 10^{-2}$ | $.36157 \times 10^{-1}$ | .00000 | .00000 |
| $S'_t$ | $-.10606 \times 10^{-2}$ | $.21418 \times 10^{-1}$ | $.16363 \times 10^{-3}$ | $.84820 \times 10^{-2}$ |
| $S'_t/S'$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ |
| $\nabla(\bar{v}T')$ | $.33501 \times 10^{-6}$ | $-.30218 \times 10^{-3}$ | $.34168 \times 10^{-5}$ | $.17712 \times 10^{-3}$ |
| $\nabla(v'\bar{T})$ | $-.60416 \times 10^{-3}$ | $.20139 \times 10^{-1}$ | .00000 | .00000 |
| $-\gamma T'$ | $.59021 \times 10^{-3}$ | $-.19389 \times 10^{-1}$ | .00000 | .00000 |
| $T'_t$ | $-.13514 \times 10^{-4}$ | $.44723 \times 10^{-3}$ | $.34168 \times 10^{-5}$ | $.17712 \times 10^{-3}$ |
| $T'_t/T'$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ | $.23066 \times 10^{-1}$ |

perature field is no longer consistent with the new circulation, and it has to adjust to the new state. As our analysis does not assume that the temperature is fixed, we can examine the mechanism by which the growth of the temperature perturbation is achieved. This mechanism is demonstrated in Table 2.

As for the salinity, the dominant destabilizing term in the surface boxes is the advection of the mean temperature gradients by the perturbation velocity. Note, however, that unlike the salinity field, the temperature is somewhat passive here because the perturbation velocity is determined by the salinity perturbation alone. This can be seen as follows: The perturbation velocity is determined from the perturbation density field through the second equation in (4), which has the same form for the mean and perturbation fields. Now, calculating the perturbation density using the perturbation temperature alone (by setting $\beta = 0$), using the perturbation salinity alone ($\alpha = 0$), and using both the perturbation salinity and temperature, we can calculate the contribution of the salinity and temperature to the perturbation velocity $v'$. Table 2 shows the perturbation velocity calculated from $S'$, $T'$, and from both ($\rho'$), and it is clear that $v'$ is determined by the salinity perturbation. The deeper water instability mechanism for the temperature is, as for the salinity, due to the advection of the perturbation temperature by the mean velocity.

We conclude that the instability of the temperature field exists only when the salinity perturbations are unstable, and the mechanism of instability for the temperature relies on the perturbation velocity created by the salinity perturbation.

Next, consider the instability regime where the unstable mode is oscillatory, that is, where the unstable eigenvalue has a nonvanishing imaginary part (regime marked "O" in Fig. 4a). In this regime, there are two eigenmodes for which both the eigenvalue and eigenvector are the complex conjugates of each other. It seems worthwhile to make sure that the instability mechanism is the one discussed above even when there are two different unstable eigenvalues. At any given moment the perturbation may be either increasing or decreasing depending on the timing within the oscillation period. Averaging the time tendencies $S'_t$ and $T'_t$ over the oscillation period given by $2\pi/\lambda^{\mathrm{Im}}$, we obtain the averaged tendency that is affected by the exponential growth or decay alone, and in which we are interested.

Let the two unstable complex eigenvectors and eigenvalues be denoted by ($\mathbf{X} = \mathbf{X}^{\mathrm{Re}} + i\mathbf{X}^{\mathrm{Im}}$, $\mathbf{X}^* = \mathbf{X}^{\mathrm{Re}} - i\mathbf{X}^{\mathrm{Im}}$) and ($\lambda$, $\lambda^*$), respectively, where $\mathbf{X}^*$ denotes the complex conjugate of $\mathbf{X}$, etc. Suppose that at time $t = 0$ the initial conditions for the temperature and salinity perturbations are such that

$$\begin{pmatrix} T' \\ S' \end{pmatrix} = \mathbf{X}^{\mathrm{Re}} = \frac{1}{2}(\mathbf{X} + \mathbf{X}^*). \tag{10}$$

At later times,

$$\begin{pmatrix} T' \\ S' \end{pmatrix} = \frac{1}{2}(\mathbf{X}e^{\lambda t} + \mathbf{X}^* e^{\lambda^* t}), \tag{11}$$

and the averaged time tendency is given by

TABLE 3. Summary of box model experiments used to demonstrate the stabilizing effect of increased salinity restoring times. Solutions ($a$, $b$, $c$) are found using restoring boundary conditions for both the temperature and salinity. Solutions ($a1$, $b1$, $c1$) are found by diagnosing the air–sea freshwater flux from solutions ($a$, $b$, $c$), and searching for the second temperature-dominated steady solution to the model equations under flux salt conditions. The rms deviations of temperature and salinity ($\Delta$SST and $\Delta$SSS, not to be confused with $\Delta S$ and $\Delta T$, which denote the equator-to-pole surface differences in temperature and salinity in the box model) are given in units of degrees Celsius and ppt; heat and freshwater flux are given in units of watts per square meter and centimeters per year; and the meridional circulation is given in Sverdrups (Sv $\equiv 10^6$ m$^3$ s$^{-1}$), as calculated from the meridional velocity by assuming that the zonal extent of the box model is equal to $L = 5000$ km.

| Run | $\frac{1}{\gamma_T}$ | $\frac{1}{\gamma_S}$ | $\Delta$SST (rms) | $\Delta$SSS (rms) | Flux | | Meridional circulation (Sv) | Mixed b.c. |
|---|---|---|---|---|---|---|---|---|
| | | | | | Heat | Water | | |
| a | 30 | 30 | .67 | 0.05 | 66.5 | 111.5 | 15.9 | unstable |
| a1 | 30 | — | .37 | 1.09 | 37.0 | 111.5 | 8.8 | unstable |
| b | 30 | 120 | .70 | 0.21 | 69.4 | 107.1 | 16.6 | stable |
| b1 | 30 | — | .34 | 1.21 | 34.1 | 107.1 | 8.0 | unstable |
| c | 120 | 30 | 2.23 | 0.05 | 55.5 | 100.0 | 14.2 | stable |
| c1 | 120 | — | 1.30 | 1.07 | 32.2 | 100.0 | 7.9 | unstable |

$$\left\langle \frac{\partial}{\partial t}\begin{pmatrix} T' \\ S' \end{pmatrix} \right\rangle = \int_0^{2\pi/\lambda^{\mathrm{Im}}} dt \frac{\partial}{\partial t}\begin{pmatrix} T' \\ S' \end{pmatrix}$$

$$= \left\langle \frac{1}{2}\mathbf{A}(\mathbf{X}e^{\lambda t} + \mathbf{X}^* e^{\lambda^* t}) \right\rangle$$

$$= \mathbf{A}\Re\{\langle \mathbf{X}e^{\lambda t}\rangle\}$$

$$= \mathbf{A}\Re\left\{\frac{\lambda^*}{|\lambda|^2}\mathbf{X}e^{\lambda t}\Big|_0^{2\pi/\lambda^{\mathrm{Im}}}\right\}$$

$$= \mathbf{A}\Re\{\lambda^*\mathbf{X}\}\frac{1}{|\lambda|^2}(e^{2\pi\lambda^{\mathrm{Re}}/\lambda^{\mathrm{Im}}} - 1). \quad (12)$$

This implies that in order to isolate the effect of the exponential growth or decay of a perturbation that is equal to $\mathbf{X}^{\mathrm{Re}}$ at $t = 0$, we should examine the different terms in the perturbation equations for a perturbation that is equal to $\Re\{\lambda^*\mathbf{X}\}$. Similarly, if the perturbation is equal to $\mathbf{X}^{\mathrm{Im}}$ at $t = 0$, the terms in the equations need to be evaluated for a perturbation that is equal to $\Im\{\lambda^*\mathbf{X}\}$.

Examining the different terms in the averaged perturbation equations following the above procedure, we find that the basic instability mechanism is roughly the same as in the nonoscillatory unstable regime: Walin's mechanism [i.e., $\nabla(v'\bar{S})$] is dominant in the upper boxes, and the second mechanism [i.e., $\nabla(\bar{v}S')$] acts in the deeper boxes. [There are some minor differences in the instability mechanism of the complex modes, where one of the two unstable modes is characterized by the vanishing of the term $\nabla(v'\bar{S})$ in one of the surface boxes. But the second unstable mode still acts as before, and in any case, such details are beyond the scope of this work.]

Having explored both the stability regimes and stability mechanisms of the thermohaline circulation in this highly simplified model, let us examine the effect of the temperature and salinity restoring coefficients on the linear stability of the solution obtained under restoring b.c.

## 4. The stabilizing effect of increased restoring times

We have seen in section 2 that when the salinity restoring time is changed from 30 days to 120 days, the steady solution of the global GCM obtained under restoring conditions remains stable upon transition to mixed boundary conditions. In the previous section (3), it was shown that there are different regimes of instability, and in particular that by changing the air–sea flux of fresh water (or the equivalent salt flux), the model may make a transition from a regime that is unstable under mixed conditions to a stable regime (see the boundary between the markers * and O in Fig. 4a).

These two observations immediately suggest a possible explanation for the stabilization of the general circulation model solution due to the increase in salinity restoring times. It seems possible that the change in salinity restoring time from 30 to 120 days caused a corresponding change in the implied surface salt flux, which in turn caused the general circulation model to make a transition to a stable regime. The purpose of this section is to carefully examine this hypothesis.

We begin by repeating the global GCM experiments with the box model presented in the previous section. First [entry (a) in Table 3] we calculate the steady solution using restoring conditions for both temperature and salinity, with restoring times of 30 days for both. The salinity restoring boundary conditions are used with the box model, the surface salinity gradient is specified to be $S_2^* - S_1^* = 2$ ppt. The solution obtained this way is stable to small perturbations under restoring conditions, but turns out to be unstable to small perturbations under mixed conditions [this is determined, as before, by examining the eigenvalues of the matrix $\mathbf{A}$ in (6)].

Next, the implied salt flux from solution (a) is diagnosed, and the steady model equations are solved using mixed boundary conditions. As mentioned previously, there are now two physical solutions. One is

always the same solution obtained under restoring conditions, and the details of the second solution, different from (a), are given by entry (a1) in Table 3. This second solution turns out again to be unstable to small perturbations. Table 3 gives the value of the salt flux for these two solutions (translated into dimensional equivalent freshwater flux), which is 111.5 cm $yr^{-1}$. Examining Fig. 4a, we see that the solutions obtained in (a, a1) are indeed both in the unstable regime.

In the next experiment with the box model (b), we increase the salinity restoring time to 120 days, and calculate the steady solution. This time the steady solution is stable to infinitesimal perturbations under both restoring and mixed boundary conditions. Diagnosing the salt flux and calculating the second solution under mixed condition (b1), we find that the second solution is unstable. The value of the salt flux obtained in these experiments [entries (b, b1) in Table 3] is 107.1 cm $yr^{-1}$, for which there is one stable and one unstable solution as seen in Fig. 4a.

Clearly the box model supports our assumption that increasing the restoring times results in a shift of the model from an unstable to a stable regime. This shift is obtained in the box model through a decrease in the implied freshwater flux, and a corresponding change to the interior steady solution. But is this also the case for the global GCM? Obviously, we cannot prove this in a rigorous manner, but there are some clear indications that this mechanism is also at work in the realistic primitive equations model.

Let us first examine the dependence of the freshwater forcing in the GCM solution on the salinity restoring coefficient. Examining the zonally averaged implied salt flux in the global GCM for runs (a) and (b) in Table 1, as given in Fig. 5, we see that the amplitude of the flux (or, more accurately, of the peaks in the zonally averaged figure, and therefore of the north–south flux gradient) is consistently larger for the 30 days restoring time. This is also the case for the box model as can be seen in Table 3 by comparing the water flux for entries (a) and (b) (111.5 vs 107.1 cm $yr^{-1}$). We have seen that in the box model the small weakening of the freshwater forcing was sufficient to shift the model to the stable regime, and it seems that this is also the case for the GCM. Note that the change in freshwater flux between the two GCM solutions, as given in Fig. 5, is not large. (The somewhat larger differences in zonally averaged fluxes at 80°S and 60°N reflects a small contribution to the total difference in the freshwater forcing, as the ocean area at these latitudes is rather small.)

Having confirmed the dependence of the freshwater forcing on the salinity restoring coefficient, let us now examine how the interior solution depends on the restoring coefficients for both the box model and the realistic GCM. Consider the meridional circulation calculated in the two GCM experiments shown in Table 1. Increasing the salinity restoring time from 30 to 120

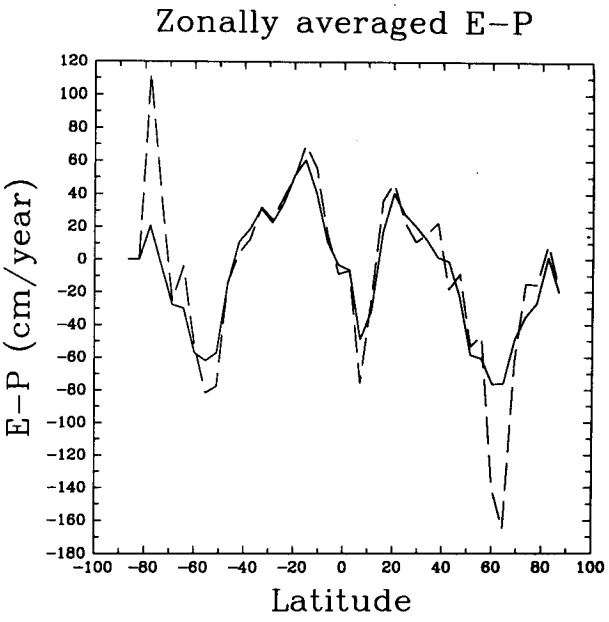



FIG. 5. Zonally averaged air–sea freshwater flux, calculated for the global primitive equations GCM under restoring conditions [entries (a) and (b) in Table 1]. The full curve is for 120-day salinity restoring time, and the dashed curve for 30-day restoring time.

days resulted in an increase of the meridional circulation in the North Atlantic Ocean from 18.3 to 22.3 Sv (Sv $\equiv 10^6$ $m^3$ $s^{-1}$). The corresponding increase in the salinity restoring time for the box model also results in an increase of the meridional circulation, as seen in Table 3 by comparing entries (a) and (b). In both cases the reason for the increase in the meridional circulation is as follows. The salinity forcing induces circulation that is opposite to that of the temperature-dominated thermohaline circulation, and acts as a break on the thermohaline circulation (Walin 1985; Marotzke et al. 1988). The increased restoring time for salinity results in weaker freshwater forcing, and therefore the braking effect of the salinity is weakened, allowing the temperature-dominated meridional circulation to strengthen. This effect seems to work in a similar way for both the GCM and the box model, indicating that the GCM is in a similar parameter regime to that of the box model.

The above diagnostics, using the meridional streamfunction and the implied air–sea salt flux, strengthen the similarity between the behavior of the box model and the global GCM as far as the response to an increased salinity restoring time is concerned. This similarity, in turn, strengthens our conclusion that the mechanism that resulted in the stabilization of the GCM as the restoring time was increased is the same mechanism we observed in the box model, that is, the transition from an unstable to a stable regime because of a change in the implied salt flux.

The close resemblance between the box model results and the GCM results is obviously not a coincidence,

but a result of fine-tuning the box model parameters. But the basic mechanism demonstrated by the box model does seem to be at work for the GCM as well. A sensitivity analysis of the box model results shows that the extent and location of the stability regimes shown in Fig. 4a are both affected by the various model parameters, such as $\Delta$, $\delta$, $S_2^* - S_1^*$, etc. In particular, the magnitude of the exponential growth time, determined by the real part of the unstable eigenvalue, can change from tens of years to thousands of years. The possible existence of very different instability time scales in the box model teaches us something about the GCM as well: The GCM solution derived using $\gamma_S^{-1} = 120$ days and which seemed stable under mixed boundary conditions may in fact be weakly unstable with a very long instability time scale. But it seems reasonable to assume that in this case a further increase of the restoring time will bring the solution into the stable regime. In any case, the values of the restoring coefficients used in Table 1 were chosen only to demonstrate the effects of increased restoring time. None of them is meant to be the optimal value to be used for this or other realistic general circulation model (especially as the flux obtained from a given restoring coefficient also depends on the vertical resolution of the model, see (1)].

The suggested stabilizing mechanism implies that the global GCM solution, obtained with $\gamma_S^{-1}$ of 30 or 120 days, is near the transition point from the stable to the unstable regime. This is deduced from the fact that the stable and unstable GCM solutions for the temperature, salinity, and circulation are very similar. In fact, the interior temperature and salinity fields for the two runs shown in Table 1 are hardly distinguishable in most areas. The proximity to the transition point allows the model to make the transition to the stable regime induced by a relatively minor change in the implied surface freshwater flux. As we believe that the ocean model solution is not far from the state of the ocean itself, this also implies that the ocean itself is near this transition point. This possibility was, in fact, already raised by Walin (1985). The proximity to the transition point makes sense physically, because as the freshwater forcing increases, the ocean goes into the unstable regime, and has to readjust the thermohaline circulation until it is back in the stable regime not far from the transition point, when the adjustment may stop.

Having understood how a change in restoring times may result in a stabilization of the solution obtained with restoring conditions, we now wish to examine how one should go about choosing a value for the restoring coefficients for realistic models.

## 5. A consistency condition for the magnitude of the restoring time

Numerical general circulation models normally contain a number of parameters that are specified fairly arbitrarily. The most outstanding example is, perhaps, the various eddy mixing coefficients, while the restoring times used in the upper boundary condition formulation are another important example. One would naturally like to reduce the uncertainty in the values of these parameters especially if the model results are sensitive to them. We have shown above that the behavior of the solution of a global primitive equation model critically depends on the value of the restoring time used for the salinity upper boundary condition. In this section we wish to propose a criterion for choosing the value of the restoring times.

When restoring coefficients are discussed, Haney's (1971) work is sometimes used to deduce the value of the restoring coefficient for the temperature. But note that Haney's formulation involves writing the air–sea heat flux as proportional to the difference between the oceanic surface temperature and an apparent atmospheric temperature, which is neither the atmospheric temperature nor the observed oceanic surface temperature. As ocean models are commonly run by restoring the surface temperature to an observed oceanic sea surface temperature, one cannot directly apply Haney's formalism to calculate the restoring coefficient for such model runs.

The best way to choose the restoring times for the salinity boundary conditions is even less clear. As there is weaker feedback between surface salinity and freshwater flux than between surface temperature and the air–sea heat flux, one tends to use larger values for the salinity restoring times than for the temperature restoring times. But the precise value to be used is still arbitrary to a large extent. The sensitivity of the model stability to the value of these coefficients, however, requires that these coefficients be chosen with care.

We may summarize our previous findings concerning the stability of the thermohaline circulation in the ocean general circulation model as follows: Consider the stability of the ocean model as function of some measure of the amplitude of the salinity forcing (either $E - P$ or $\beta\Delta S/\alpha\Delta T$, see Fig. 4a). The entire range of forcing is divided into stable and unstable regimes. It seems plausible, based on what we have seen, that the ocean itself lies in the stable regime near the stability transition point, while the steady solution of an ocean model obtained with restoring conditions may lie in the unstable regime or the stable regime, depending on the magnitude of the salinity restoring coefficient.

Suppose the ocean model dynamics were completely consistent with the real ocean's dynamics, and that the observed sea surface temperature and salinity were known with no observational errors. Then prescribing the model surface temperature and salinity to be equal exactly to the observed fields would result in the model solution being in the stable regime, near the transition point, just like the real ocean.

But there obviously are inconsistencies between the model dynamics and those of the real ocean, especially

for coarse-resolution models as used in climate studies. Similarly, there are observational errors in the measured surface temperature and salinity fields that are used to constrain ocean models. Because of these errors, the transition point for the ocean model may be in a slightly different position from that of the real ocean. Presumably, the distance between the stability transition points of the ocean and the model is of the order of the combined observational and model errors. Constraining the ocean model precisely to the observed fields may therefore result in the model solution being near the stability transition point, but in the unstable regime. Figure 6 schematically describes a possible configuration for the location of the stability transition points and the location of the GCM solutions. Note in particular that while the solution obtained with a salinity restoring time of 30 days may be characterized by surface forcing that is in the stable regime for the ocean itself, it is in the unstable regime for the model. Clearly our discussion of the reasons for the stabilization of the global PE model is speculative to some extent, as a rigorous stability analysis such as was carried out for the box model is simply not possible for the three dimensional primitive equation general circulation model.

The above explanation also suggests a remedy for the undesired instability with respect to mixed b.c.: one should not set the model surface temperature and salinity fields exactly equal to the observed fields, but allow for some deviation. This deviation should be within the limits determined by both the observational error and the error expected to result from inconsistencies between the model and oceanic dynamics. It is this deviation that allows the model to "move" into the stable regime. (While one may have some rough idea of the magnitude of the observational error, it is more difficult to quantify the model error for this purpose. It may roughly be defined, perhaps, to be of the order of magnitude of the difference between the model results and the data.)

Note that increasing the restoring times results in larger deviations of the calculated surface properties from the observed ones, and in weaker salt forcing that may bring the model into the stable regime. We can therefore formulate a criterion for choosing the right values of the restoring coefficients: The restoring coefficients should be chosen such that at steady state, the root-mean-square deviation between the model and observed surface fields is of the order of the combined observational error and model error:

choose $\gamma_T$ and $\gamma_S$ such that

$$\left.\begin{aligned}\mathrm{rms}[\mathrm{SST}_{\mathrm{model}} - \mathrm{SST}_{\mathrm{data}}]\\ \mathrm{rms}[\mathrm{SSS}_{\mathrm{model}} - \mathrm{SSS}_{\mathrm{data}}]\end{aligned}\right\} \approx \text{observational error} + \text{model error}.$$

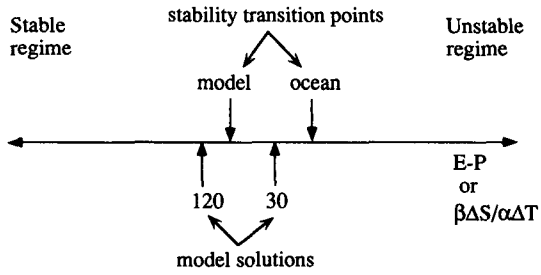



FIG. 6. A schematic picture of a possible configuration for the location of the stability transition points and the location of the GCM solutions. See text for details.

Returning now to the GCM runs in Table 1 and the box model runs in Table 3, we first note that the increased salinity restoring time resulted in a larger rms deviation of the model surface temperature and salinity from the observed fields. Using 30 days restoring time for the salinity in the global GCM resulted in an rms salinity deviation of 0.11 ppt, certainly less than even the observational error alone for the climatological annually averaged data we use. Increasing the restoring time to 120 days increased the rms salinity deviation to 0.22 ppt, which is more acceptable, and indeed resulted in a steady solution that is stable under mixed boundary conditions. The rms deviation of the GCM surface temperature from the observed surface temperature is about 0.5°C, which is probably less than the observational error. We have examined the effect of increasing the restoring time for temperature using the box model only, and found that this too can shift the box model from an unstable to a stable regime [entry (c) in Table 3].

Examining the spatial structure of the implied freshwater flux in the global GCM for the solutions shown in Figs. 1 and 2, we note that the solution with 30-day salinity restoring time is much noisier than that obtained with a salinity restoring time of 120 days. The noisy structure of $E - P$ field for $\gamma_S^{-1} = 30$ days seems more an artifact of the calculation than a real feature of the freshwater forcing in the ocean. The results with 120 days, on the other hand, seem significantly more reasonable. Examining the surface salinity and meridional salinity sections through the various oceans for the two solutions (not shown), one can see that the differences between the two are very small, within the observational error in the Levitus (1982) data.

To summarize, it seems that the increased salinity restoring time produces seemingly more reasonable results for the freshwater flux, does not cause too large deviations from the specified surface salinity, and most important, results in a solution that seems stable upon transition to mixed boundary conditions.

In a recent work Zhang et al. (1993) argued that the formulation of mixed boundary conditions with a relatively short restoring time for the temperature is inconsistent with the small atmospheric heat capacity

relative to that of the upper ocean. They argued that significantly longer restoring times should be used, and showed that this may result in different behavior under mixed boundary conditions. Their work seems to indicate that the polar halocline catastrophe of Bryan (1986), which is intimately related to the instability under mixed b.c., may not be as severe a problem for climate models as it may seem at first. Thus, they support the conclusion of this study, although from a completely different point of view than ours.

## 6. Conclusions

We have used a global primitive equations oceanic general circulation model and a simple four-box model of the meridional circulation to examine and analyze the stability of the thermohaline circulation in a realistic ocean model under mixed boundary conditions. This instability has been extensively documented using models of various complexities, but within idealized geometries. Our purpose was to examine this instability for a GCM that is run with realistic geometry and forcing conditions, as is done in climate studies. More specifically, we tried to determine whether the thermohaline instability should occur in the parameter range characterizing today's oceans.

We found that for a realistic GCM, using realistic geometry and forcing there are both stable and unstable regimes with respect to mixed boundary conditions. The steady solution of the global GCM under restoring conditions may be either stable or unstable upon transition to mixed boundary conditions, depending on the regime in which the solution is. One of the most important findings here is that the solution of a general circulation model using realistic geometry and forcing lies in the stable regime, but very close to the stability transition point with respect to mixed boundary conditions. This proximity to the transition point allows the model to make a transition between the unstable and stable regimes due to a relatively minor change in the surface freshwater flux and in the interior solution. Such a change in the surface flux may be induced, for example, by changing the salinity restoring time used to obtain the steady model solution under restoring conditions. It is important to note that the needed change in the restoring time in order to shift the model from one stability regime to another is not large, and the GCM solutions obtained in both regimes are nearly identical within observational errors. Assuming that the ocean model solution is not far from the state of the ocean itself, this also implies that the ocean itself is near this stability transition point, as also suggested by Walin (1985).

By carefully analyzing the stability regimes of a simple four-box model, we were able to demonstrate that an increase in the restoring time for salinity may indeed cause the box model to make a transition from an unstable regime to a stable regime. Such a rigorous stability analysis is, of course, impossible to carry out using the PE model, but we have presented and examined evidence indicating that such a transition is also responsible for the stabilizing effect of increased restoring time in the realistic three-dimensional general circulation model.

Finally, we argued that the use of too short restoring times is inconsistent with the level of errors in the data and in the model dynamics. A consistency criterion for the magnitude of the restoring times in realistic models was formulated. This criterion states that the restoring coefficients should be large enough to allow the rms deviation of the model surface temperature and salinity from the corresponding observed fields to be of the order of the observational and model errors. Shorter restoring times bring the calculated surface fields nearer to the observed ones. But because the transition point from stable to unstable regimes is different in the ocean and in the model, the more realistic surface fields seem to correspond to an unstable regime in the ocean model.

Our finding that a realistic ocean model using carefully formulated boundary conditions should be stable upon transition to mixed boundary conditions is also relevant to some recent studies concerning the variability of the oceanic circulation. Weaver and Sarachik (1991) have shown that with strong freshwater forcing, a primitive equations model may produce strong interdecadal variability. They also showed that this variability does not exist for weaker freshwater forcing, and that the parameter range where this variability occurs is within the unstable regime with respect to mixed boundary conditions. Our results, on the other hand, seem to indicate that the present day oceans are in a stable regime, or at least very close to this regime and not in a strong forcing unstable regime. It is still possible, of course, that a model of higher resolution and with lower viscosity and diffusion coefficients, may be able to excite the decadal oscillations mechanism found by Weaver et al. (1991) even in the seemingly weaker freshwater forcing that corresponds to the state of today's oceans.

Climate studies often initialize a coupled ocean–atmosphere model run with the steady-state oceanic circulation obtained with restoring boundary conditions using climatological surface temperature and salinity. The results presented here may be of use to such studies by indicating how the restoring coefficients for the initialization run should be chosen in order to avoid the instability upon transition to the coupled model run. The solutions found using the longer restoring times are sufficiently close to the observed temperature and salinity fields, like those found with shorter restoring times. Yet we have seen that the implied surface fluxes are far smoother and more reasonable. Coupled models are often run with a "flux correction" (Manabe et al. 1991) calculated from the difference between the implied fluxes from the ocean model and those from the

atmospheric model. The more reasonable surface fluxes obtained from the ocean model using carefully chosen restoring coefficients may significantly reduce the flux correction needed to prevent a climate drift in the coupled model.

*Acknowledgments.* This work began during a visit of ET to the Geophysical Fluid Dynamics Laboratory. We are grateful to Bonnie Samuels and Keith Dixon for their kind help. Comments by Michael Ghil and an anonymous reviewer were most useful in clarifying the presentation.

## REFERENCES

Bryan, F., 1986: High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature,* **323**(25), 301–304.

Bryan, K., 1969: A numerical method for the study of the circulation of the world ocean. *J. Comput. Phys.,* **4,** 347–376.

Cox, M. D., 1984: A primitive equation 3 dimensional model of the ocean. GFDL ocean group Tech. rep. No 1. Princeton University.

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

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

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

Levitus, S., 1982: *Climatological Atlas of the World Ocean.* NOAA Tech. Paper 3, 173 pp.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, 1991: Transient response of a coupled ocean–atmosphere model to gradual changes of atmospheric $CO_2$. Part I: Annual mean response. *J. Climate,* **4,** 785–818.

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

——, 1990: Instability and Multiple Equilibria of the thermohaline circulation. *Ber. Inst. Meereskd. Univ. Kiel,* **194,** 126 pp.

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

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

Numerical Algorithms Group, 1984: Fortran Library Manual, Mark II, 6 vols.

Quon, C., and M. Ghil, 1993: Multiple equilibria in thermosolutal convection due to salt-flux boundary conditions. *J. Fluid Mech.,* in press.

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

Semtner, A. J., 1974: An oceanic general circulation model with bottom topography. UCLA Dept. of Meteorology Tech. Rep. No. 9, 99 pp.

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

Stouffer, R. J., S. Manabe, and K. Bryan, 1989: Interhemispheric asymmetry in climate response to a gradual increase of atmospheric $CO_2$. *Nature,* **342,** 660–662.

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

Toggweiler, J. R., and B. Samuels, 1993: Is the magnitude of the deep outflow from the Atlantic Ocean actually governed by Southern Hemisphere winds? *The Global Carbon Cycle,* M. Heimann, Ed., NATO ASI Series, Springer-Verlag, 303–331.

Walin, G., 1985: The thermohaline circulation and the control of ice ages. *Paleaogeogr. Paleoclim., Paleoecol.,* **50,** 323–332.

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.

——, ——, and J. Marotzke, 1991: Freshwater flux forcing of decadal and interdecadal oceanic variability. *Nature,* **353,** 836–838.

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