Thermohaline Circulation Stability: A Box Model Study. Part I: Uncoupled Model

VALERIO LUCARINI* AND PETER H. STONE

Department of Earth, Atmospheric and Planetary Sciences, and Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 23 June 2003, in final form 29 July 2004)

ABSTRACT

A thorough analysis of the stability of the uncoupled Rooth interhemispheric three-box model of thermohaline circulation (THC) is presented. The model consists of a northern high-latitude box, a tropical box, and a southern high-latitude box, which correspond to the northern, tropical, and southern Atlantic Ocean, respectively. Restoring boundary conditions are adopted for the temperature variables, and flux boundary conditions are adopted for the salinity variables. This paper examines how the strength of THC changes when the system undergoes forcings that are analogous to those of global warming conditions by applying the equilibrium state perturbations to the moisture and heat fluxes into the three boxes. In each class of experiments, using suitably defined metrics, the authors determine the boundary dividing the set of forcing scenarios that lead the system to equilibria characterized by a THC pattern similar to the present one from those that drive the system to equilibria with a reversed THC. Fast increases in the moisture flux into the northern high-latitude box are more effective than slow increases in leading the THC to a breakdown, while the increases of moisture flux into the southern high-latitude box strongly inhibit the breakdown and can prevent it, as in the case of slow increases in the Northern Hemisphere. High rates of heat flux increase in the Northern Hemisphere destabilize the system more effectively than low ones; increases in the heat fluxes in the Southern Hemisphere tend to stabilize the system.

1. Introduction

Following the classical picture of Sandström's (Sandström 1908; Huang 1999), ocean circulation can be divided, albeit somewhat ambiguously, into the wind-driven circulation, which is directly induced by the mechanical action of the wind and essentially affects the surface waters, and the thermohaline circulation (THC), which is characterized by relatively robust gradients in the buoyancy of the water masses (Weaver and Hughes 1992; Rahmstorf 2000, 2002; Stocker et al. 2001). Although some studies show that the THC is very sensitive to wind stresses (Toggweiler and Samuels 1995), more recent studies suggest that the buoyancy forces are dominant when more realistic boundary conditions are considered (Rahmstorf and England 1997; Bugnion and Hill 2003). The THC dominates the large-scale picture of global ocean circulation and is especially relevant for intermediate-to-deep water.

One of the main issues in the study of climate change is the fate of the THC in the context of global warming (Weaver and Hughes 1992; Manabe and Stouffer 1993; Stocker and Schmittner 1997; Rahmstorf 1997, 1999a,b, 2000; Wang et al. 1999a,b). The THC plays a major role in the global circulation of the oceans as pictured by the conveyor belt scheme (Weaver and Hughes 1992; Stocker 2001); therefore, the effect on the climate of a change of its pattern may be global (Broecker 1997; Manabe and Stouffer 1999b; Cubasch et al. 2001; Seager et al. 2002). The THC is sensitive to changes in the climate since the North Atlantic Deep Water (NADW) formation is sensitive to variations in air temperature and precipitation in the Atlantic basin (Rahmstorf and Willebrand 1995; Rahmstorf 1996). There are several paleoclimatic datasets indicating that changes in the patterns or collapses of the THC coincide with large variations in climate, especially in the North Atlantic region (Broecker et al. 1985; Boyle and Keigwin 1987; Keigwin et al. 1994; Rahmstorf 2002).

Box models have historically played a major role in understanding the fundamental dynamics of the THC (Weaver and Hughes 1992). The Stommel two-box oceanic model (Stommel 1961) parameterized the THC

* Current affiliation: Department of Mathematics and Computer Science, University of Camerino, Camerino, Italy.

Corresponding author address: Dr. Valerio Lucarini, Dept. of Mathematics and Computer Science, University of Camerino, via Madonna delle Carceri 62032, Camerino, Italy.
E-mail: lucarini@alum.mit.edu

strength as proportional to the density gradient between the northern latitude and equatorial boxes. The oceanic model proposed by Rooth (1982) introduced the idea that the driver of the THC was the density difference between the high-latitude basins of the Northern and Southern Hemispheres, thus implying that the THC is an interhemispheric phenomenon. Both the Stommel and Rooth models allow two stable equilibria, one—the warm mode—characterized by the downwelling of water in the high latitudes of the North Atlantic, which resembles the present oceanic circulation, and the other—the cold mode—characterized by the upwelling of water in the high latitudes of the North Atlantic. Perturbations to the driving parameters of the system, that is, freshwater and heat fluxes in the oceanic boxes, can cause transitions between the two regimes.

Various GCM experiments have shown that multiple equilibria of the THC are possible (Bryan 1986; Manabe and Stouffer 1988; Stouffer et al. 1991; Stocker and Wright 1991; Manabe and Stouffer 1999a; Marotzke and Willebrand 1991; Hughes and Weaver 1994), and most GCMs have shown that the greenhouse gases (GHGs) radiative forcing could cause a weakening of the THC by the inhibition of the sinking of the water in the northern Atlantic. Large increases of the moisture flux and/or of the surface air temperature in the downwelling regions are the driving mechanisms of such a process. Nevertheless, in a recent global warming simulation run the THC remained almost constant since the destabilizing mechanisms have been offset by the increase of the salinity advected into the downwelling region due to net freshwater export from the whole North Atlantic (Latif et al. 2000).

Hemispheric box models along the lines of Stommel's have been extensively studied, and the stability of their THC has been assessed in the context of various levels of complexity in the representation of the coupling between the ocean and the atmosphere (Weaver and Hughes 1992; Nakamura et al. 1994; Marotzke 1996; Krasovskiy and Stone 1998; Tziperman and Gildor 2002).

However, analysis of GCM data has indicated that the THC is an interhemispheric phenomenon, and in some cases it has been found that the THC strength is approximatively proportional to the density difference between the northern and southern Atlantic (Hughes and Weaver 1994; Rahmstorf 1996; Klinger and Marotzke 1999; Wang et al. 1999a), thus supporting Rooth's approach.

In this paper we perform an analysis of the stability of the present pattern of the THC against changes to the heat and freshwater atmospheric fluxes using an uncoupled version of the Rooth oceanic model. We note that this model assumes that buoyancy is well mixed in the vertical, and thus avoids the need for parameterizing the mixing (Munk and Wunsch 1998). The model considered here has no real explicit coupling between the ocean and the atmosphere: the atmospheric freshwater fluxes are prescribed and the atmospheric heat fluxes relax the oceanic temperatures toward prescribed values (Rahmstorf 1996; Scott et al. 1999; Stone and Krasovskiy 1999; Titz et al. 2002a,b).

We explicitly analyze the role of the spatial patterns and the rates of increase of the forcings in determining the response of the system. We consider a suitably defined metric and determine what are the thresholds beyond which there is destabilization of the warm mode of the THC. We emphasize that our treatment goes beyond quasi-static analysis since the effect of the rate of forcing is explicitly addressed, so we analyze how the effects on the system of slow (Wang et al. 1999a) and rapid (Wiebe and Weaver 1999) changes contribute. In particular, in the limit of very slow changes, our results coincide with those that can be obtained with the analysis of the bifurcations of the system (Stone and Krasovskiy 1999; Scott et al. 1999). Relatively few studies with more complex models have explicitly addressed how the spatial (Rahmstorf 1995, 1996, 1997) or temporal (Stocker and Schmittner 1997) patterns of forcing determine the response of the system. However, these studies did not carry out a parametric study of the influence of both spatial and temporal patterns. We emphasize that our study is based on a process model that aims more at obtaining a qualitative picture of the properties of the THC than providing quantitative results.

Our paper is organized as follows: in section 2 we provide a description and the general mathematical formulation of the dynamics of the three-box model used in this study. In section 3 we describe the parameterization of the physical processes. In section 4 we analyze the feedbacks of the system. In section 5 we describe the stability properties of the system when it is forced with perturbations in the freshwater flux. In section 6 we describe the stability properties of the system when it is forced with perturbations in the heat flux. In section 7 we present our conclusions.

2. Model description

The three-box model consists of a northern high-latitude box (box 1), a tropical box (box 2), and a southern high-latitude box (box 3). The volume of the two high-latitude boxes is the same and is $1/V$ times the volume of the tropical box. We choose $V = 2$ so that box 1, box 2, and box 3 can be thought of as describing the portions of an ocean like the Atlantic north of 30°N, between 30°N and 30°S, and south of 30°S, respectively. At the eastern and western boundaries of the oceanic boxes there is land; our oceanic system spans 60° in longitude so that it covers $\epsilon = 1/6$ of the total surface of the planet. The boxes are 5000 m deep, so that the total mass $M_{i=1,3} = M$ of the water contained in each of the high-latitude boxes is $\approx 1.1 \times 10^{20}$ kg, while $M_2 = V \times M_{i=1,3} = 2 \times M$. The tropical box is connected to both high-latitude boxes; moreover, the two high-latitude boxes are connected by a deep current passage (which bypasses the tropical box) containing a negligible mass. The three boxes are assumed to be well mixed. The assumption of well-mixed boxes excludes the polar halocline disaster (Bryan 1986; Weaver and Hughes 1992; Zhang et al. 1993) from the range of phenomena that can be described by this model. In our model we have not included a deep box as done in Rahmstorf (1996). This choice allows an easier and more detailed control of the behavior of the system, but implies that distortions are introduced in the response of the system to fast perturbations (Weaver and Hughes 1992). The physical state of the box i is described by its temperature T_i and its salinity S_i; the box i receives from the atmosphere the net flux of heat $\tilde{H}_i$ and the net flux of freshwater $\tilde{F}_i$; the freshwater fluxes $\tilde{F}_i$ globally sum up to 0, since the freshwater is a globally conserved quantity. The box i is subjected to the oceanic advection of heat and salt from the upstream box through the THC, whose strength is $\tilde{q}$.

In Fig. 1 we present a scheme of our system in the northern sinking pattern: note that the arrows representative of the freshwater fluxes are arranged in such a way that the conservation law is automatically included in the graph. The dynamics of the system is described by the tendency equations for the heat and the salinity of each box. We divide the three heat tendency equations by $c_p \times M_i$, where c_p is the constant pressure specific heat of water per unit of mass, and in the salinity tendency equations we neglect the contribution of the freshwater fluxes in the mass balance, so that virtual salinity fluxes and freshwater fluxes are equivalent (Marotzke 1996; Rahmstorf 1996; Scott et al. 1999). We obtain the following final form for the temperature and salinity tendency equations for the three boxes (Scott et al. 1999):

$$\dot{T}_1 = \begin{cases} q(T_2 - T_1) + H_1, & q \geq 0 \\ q(T_1 - T_3) + H_1, & q < 0 \end{cases}, \tag{1}$$

$$\dot{T}_2 = \begin{cases} \dfrac{q}{V}(T_3 - T_2) + H_2, & q \geq 0 \\ \dfrac{q}{V}(T_2 - T_1) + H_2, & q < 0 \end{cases}, \tag{2}$$

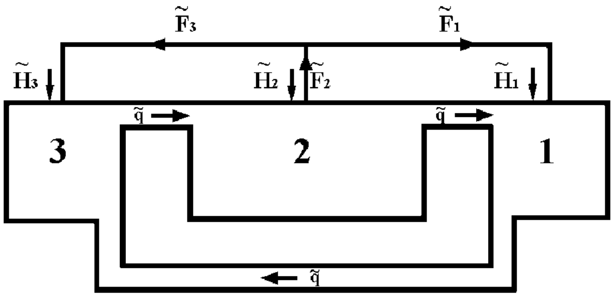

FIG. 1. Schematic picture of the interhemispheric box model.

$$\dot{T}_3 = \begin{cases} q(T_1 - T_3) + H_3, & q \geq 0 \\ q(T_3 - T_2) + H_3, & q < 0 \end{cases}, \tag{3}$$

$$\dot{S}_1 = \begin{cases} q(S_2 - S_1) - F_1, & q \geq 0 \\ q(S_1 - S_3) - F_1, & q < 0 \end{cases}, \tag{4}$$

$$\dot{S}_2 = \begin{cases} \dfrac{q}{V}(S_3 - S_2) - F_2, & q \geq 0 \\ \dfrac{q}{V}(S_2 - S_1) - F_2, & q < 0 \end{cases}, \tag{5}$$

$$\dot{S}_3 = \begin{cases} q(S_1 - S_3) - F_3, & q \geq 0 \\ q(S_3 - S_2) - F_3, & q < 0 \end{cases}, \tag{6}$$

where $q = \rho_0\tilde{q}/M$, $H_i = \tilde{H}_i/(c_pM)$, and $F_i = \rho_0 S_0 \tilde{F}_i/M$.

We impose the condition that the average salinity is a conserved quantity; this means that $\dot{S}_1 + V\dot{S}_2 + \dot{S}_3 = 0$, which implies that

$$F_2 = -\frac{1}{V}(F_1 + F_3). \tag{7}$$

This conservation law holds at every time, and so rules out the possibility of including the effects of the melting of continental ice sheets in our study. We note that the system can (asymptotically) reach an equilibrium only if its feedbacks can drive it toward a state in which the following equation holds:

$$H_2 = -\frac{1}{V}(H_1 + H_3). \tag{8}$$

The strength of the specific THC is parameterized as proportional to the difference between the density of the water contained in the box 1 and the density of the water contained in the box 3, as in the Rooth (1982) model. Given that the water density can approximately be expressed as

$$\rho(T, S) \approx \rho_0(1 - \alpha T + \beta S), \tag{9}$$

where α and β are the thermal and haline expansion coefficients, respectively; set to 1.5×10^{-4} °C^{-1} and 8×10^{-4} psu^{-1}, respectively (where psu stands for practical salinity unit), and ρ_0 is a standard reference density, we obtain for the normalized THC strength q the following relation:

$$q = \frac{k}{\rho_0}(\rho_1 - \rho_3) = k[\alpha(T_3 - T_1) + \beta(S_1 - S_3)], \tag{10}$$

where k is the hydraulic constant, chosen to obtain a reasonable value of the THC strength. We emphasize that the constant k roughly summarizes the wind-powered turbulent mixing taking place in the ocean and powering the THC, and we think that, in principle, it should be expressed as a function of the climatic state of the system rather than being a constant.

The northern (southern) sinking state of the circulation is therefore characterized by $q > (<)0$. Considering the case $q > 0$, we obtain a diagnostic relation for steady state q by setting to zero Eqs. (3) and (6), and using Eq. (9):

$$q^{eq} = [k(\alpha H_3^{eq} + \beta F_3^{eq})]^{1/2}; \tag{11}$$

in the case $q < 0$, a diagnostic relation for the steady state q can be obtained using the same procedure as above but setting to zero Eqs. (1) and (4):

$$q^{eq} = -[k(\alpha H_1^{eq} + \beta F_1^{eq})]^{1/2}. \tag{12}$$

We conclude that the equilibrium value of the THC strength is determined by the equilibrium values of the heat and freshwater fluxes of the box where we have upwelling of water. These results generalize the expressions given in Rahmstorf (1996). The transient evolution of q is determined by its tendency equation:

$$\dot{q} = -q^2 + kq[\alpha(T_1 - T_2) - \beta(S_1 - S_2)] + \\ + k[\alpha(H_3 - H_1) + \beta(F_3 - F_1)], \quad q \geq 0, \tag{13}$$

$$\dot{q} = q^2 + kq[\alpha(T_3 - T_2) - \beta(S_3 - S_2)] + \\ + k[\alpha(H_1 - H_3) + \beta(F_1 - F_3)], \quad q < 0. \tag{14}$$

We observe that the difference between the forcings applied to the freshwater and heat fluxes into boxes 1 and 3 directly affect the evolution of q; the presence of terms involving the gradient of temperature and salinity between box 2 and box 1 (box 3) if $q > 0$ ($q < 0$) breaks the symmetry of the role played by the two high-latitude boxes. In our experiments we perturb an initial equilibrium state by changing at various rates the parameters controlling the fluxes H_i and/or F_i, and observe under which conditions we obtain a reversal of the THC. The reversal of the THC causes a sudden cooling and freshening in box 1 and a sudden warming and increase of the salinity in box 3, because the former loses and the latter receives the advection of the tropical warm and salty water.

3. Parameterization of the physical processes

We perform our study considering the original formulation of the Rooth (1982) model and thus consider mixed boundary conditions for the surface fluxes. For a more detailed description of the model see also Scott et al. (1999). The heat flux into the box i is described by a Newtonian relaxation law of the form $H_i = \lambda_i(\tau_i - T_i)$, where the parameter τ_i is the target temperature (Bretherton 1982), which represents a climatologically reasonable value toward which the box temperature is relaxed, and the parameter λ_i quantifies the efficacy of the relaxation. Along the lines of Scott et al. (1999), and following the parameterization introduced by Marotzke (1996), we choose $\tau_1 = \tau_3 = 0°C$ and $\tau_2 = 30°C$, and select $\lambda_1 = \lambda_2 = \lambda_3 = \lambda = 1.29 \times 10^{-9}$ s^{-1}, which mimics the combined effect of radiative heat transfer and of a meridional heat transport parameterized as being linearly proportional to the meridional temperature gradient (Marotzke and Stone 1995; Marotzke 1996; Scott et al. 1999). In physical terms, λ corresponds to a restoring coefficient of $\approx$4.3 W m^{-2} °C^{-1} affecting the whole planetary surface, which translates into an effective $\tilde{\lambda} \approx 1/\epsilon \times 4.3$ Wm^{-2} °C^{-1} = 25.8 Wm^{-2} °C^{-1} relative to the oceanic surface fraction only.

Using Eqs. (1)–(3), we can obtain the following tendency equation for the globally averaged temperature $T_M = (T_1 + VT_2 + T_3)/(2 + V)$:

$$\dot{T}_M = \lambda(\tau_M - T_M), \tag{15}$$

where τ_M is the average target temperature; the average temperature relaxes toward its target value in about 25 yr, which is very close to the value considered by Titz et al. (2002b). We observe that in the case of an ocean with vertical resolution, our heat flux parameterization corresponds to a restoring time of ~3 months for a 50-m-deep surface level (Marotzke and Stone 1995).

The virtual salinity fluxes F_i are not functions of any of the state variables of the system, and are given parameters. We choose $F_1 = 13.5 \times 10^{-11}$ psu s^{-1}, and $F_3 = 9 \times 10^{-11}$ psu s^{-1}, which correspond to net atmospheric freshwater fluxes of ~0.40 Sυ (1 Sv $\equiv 10^6$ m^3 s^{-1}) into box 1 and of ~0.27 Sv into box 3, respectively. These values are based on the analysis of the hydrology of the two hemispheres presented by Baumgartner and Reichel (1975) and Peixoto and Oort (1992). Our model neglects moisture fluxes associated with the wind-driven circulation. If these were strong enough to reverse the sign of F_3, the model's behavior would by very different (Rahmstorf 1996; Titz et al. 2002b), but would not be in accord with observations.

Choosing $k = 1.5 \times 10^{-6}$ s^{-1}, at equilibrium we have $q \approx 1.47 \times 10^{-10}$ s^{-1} (Scott et al. 1999). In physical terms this value describes a THC in the northern sinking pattern with a strength of $\approx$15.5 Sv; this value agrees reasonably well with estimates coming from observations (Roemmich and Wunsch 1985; Macdonald and Wunsch 1996). We emphasize that this THC strength introduces an oceanic time scale $t_s \approx 250$ yr, which is the ratio between the volumes of box 1 and 3 and the equilibrium state water transport. We refer to this time scale as the flushing or advective time of the ocean since it is the time it takes for the ocean circulation to empty out box 1 or box 3, as well as to propagate the information from one box to another. We emphasize that, since we have considered the approximation of well-mixed boxes, our model should not be considered very reliable for phenomena taking place in time scales $\ll t_s$.

In this equilibrium state, boxes 1 and 3 receive a net poleward oceanic heat advection of ~1.58 PW and ~0.17 PW, respectively. The former agrees reasonably well with estimates. The latter agrees with the sign of the global transport in the Southern Hemisphere, but disagrees with the sign in just the southern Atlantic

(Peixoto and Oort 1992; Macdonald and Wunsch 1996). The equilibrium temperature of box 1 is larger than the equilibrium temperature of box 3, and the same inequality holds for the salinities, the main reason being that box 1 receives directly the warm and salty tropical water. Taking into account Eq. (9) and considering that the actual equilibrium q is positive, we conclude that the THC is haline dominated; the ratio between the absolute value of the thermal and of the haline contribution on the right-hand side of Eq. (9) is $\approx$0.8. The occurrence of the haline dominated mode depends of course on the forcing parameters, which we have based on observations. The haline mode, with its positive value for q, is consistent with observations. We report in Table 1 the values of the main model constants, while in Table 2 we present the fundamental parameters characterizing this equilibrium state.

4. Feedbacks of the system

The Newtonian relaxation law for the temperature implies that the atmosphere has a negative thermic feedback: T_i increases $\Rightarrow$ H_i decreases $\Rightarrow$ T_i decreases.

The reaction of the ocean to a decrease in the strength of the THC can be described as follows: q decreases $\Rightarrow$

1) T_1 decreases more than T_3 $\Rightarrow$
   a) q increases,
   b) the change is limited by the atmospheric feedback;
2) S_1 decreases more than S_3 $\Rightarrow$ q decreases.

The heat advection feedback mechanism 1 is negative and dominates the positive salinity advection feedback mechanism 2. We emphasize that the sign of the salinity feedback depends on the asymmetry in the freshwater fluxes (Scott et al. 1999). We note that the strength of the second-order feedback mechanism 1b is relevant in establishing the overall stability of the THC (Tziperman et al. 1994); a very strong temperature-restoring atmospheric feedback like that in our model tends to decrease the stability of the system (Nakamura et al. 1994; Rahmstorf 2000). The feedback mechanisms 1 + 2 are governed by q so that their time scale is around the flushing time $t_s \approx 250$ yr (Scott et al. 1999) of the oceanic boxes. We expect that a fast perturbation avoids those feedbacks, since it bypasses the ocean-controlled transfer of information needed to activate them.

In order to simulate global warming conditions, we perform two sets of experiments. In the first, we increase the freshwater fluxes into the two high-latitude boxes. The rationale for this forcing is that an increase in the global mean temperature produces an increase in the moisture capacity of the atmosphere, and this is likely to cause the enhancement of the hydrological cycle, especially in the transport of moisture from the Tropics to the high latitudes. In the second, we represent the purely thermic effects of global warming by setting the increase in the target temperatures of the two high-latitude boxes larger than that of the tropical boxes, since virtually all GCM simulations forecast a larger increase in temperatures in the high latitudes (the so-called polar amplification).

TABLE 1. Value of the main model constants.

Quantity	Symbol	Value
Mass of box $i = 1, 3$	M	1.08×10^{20} kg
Box 2/box $i = 1, 3$ mass ratio	V	2
Average water density	ρ_0	1000 kg m^{-3}
Specific heat per unit mass of water	c_p	$4 \cdot 10^3\ J\ °C^{-1}kg^{-1}$
Average salinity	S_0	35 psu
Oceanic fractional area	ϵ	1/6
Thermal expansion coefficient	α	$1.5 \cdot 10^{-4}\ °C^{-1}$
Haline expansion coefficient	β	$8 \cdot 10^{-4} psu^{-1}$
Hydraulic constant	k	$1.5 \cdot 10^{-6}\ s^{-1}$
Atmospheric freshwater flux—Box 1	$\tilde{F}_1$	0.41 Sv
Atmospheric freshwater flux—Box 2	$\tilde{F}_2$	−0.68 Sv
Atmospheric freshwater flux—Box 3	$\tilde{F}_3$	0.27 Sv
Target temperature—Box 1	τ_1	0 °C
Target temperature—Box 2	τ_2	30 °C
Target temperature—Box 3	τ_3	0 °C
Atmospheric temperature restoring coefficient*	λ	25.8 Wm^{-2} °C^{-1}

* Value relative to the oceanic surface fraction only.

TABLE 2. Value of the fundamental variables of the system at the initial equilibrium state.

Variable	Box 1	Box 2	Box 3
Temperature	2.9°C	28.4°C	0.3°C
Salinity	34.7 psu	35.6 psu	34.1 psu
Surface heat flux	−1.58 PW	1.74 PW	−0.16 PW
Oceanic heat flux	1.58 PW	−1.74 PW	0.16 PW
THC strength	15.5 Sv	15.5 Sv	15.5 Sv

5. Freshwater flux forcing

We force the previously defined equilibrium state with a net freshwater flux into the two high-latitude boxes that increase linearly with time; this is obtained by prescribing

$$F_i(t) = \begin{cases} F_i(0) + F_i^t t, & 0 \le t \le t_0 \quad i = 1, 3 \\ F_i(0) + F_i^t t_0 & t > t_0 \qquad\quad i = 1, 3 \end{cases}. \tag{16}$$

We impose the conservation of the globally averaged salinity of the oceanic system by requiring that Eq. (7) holds at all times. When the perturbation is over, we reach a final state that can be either qualitatively similar to the initial northern sinking equilibrium or radically different, that is, presenting a reversed THC. Each perturbation can be uniquely specified by a set of three parameters, such as the triplet 1 [t_0, $\Delta F_3/\Delta F_1$, ΔF_1], or the triplet 2 [t_0, $\Delta F_3/\Delta F_1$, F_1^t], where $\Delta F_i \equiv F_i^t \times t_0$ is the total change of the freshwater flux into box i. In order to describe a global warming scenario, we explore the effect of positive ΔF_1 and ΔF_3. Freshwater forcing into box 1 tends to destabilize the THC, since it induces a freshening, and so, a decrease of the density of the water. If the freshwater forcing is larger in box 3, we do not obtain a reversal of the THC, the main reason being that increases in the net freshwater flux into box 3 tend to stabilize the circulation, as can be seen in Eq. (11). Since we are interested in the destabilizing perturbations, we limit ourselves to the case $\Delta F_3/\Delta F_1 < 1$. These perturbations cause an initial decrease in the THC strength q and so trigger the following feedbacks: F_1's increase is larger than F_3's $\Rightarrow S_1$ decreases more than $S_3 \Rightarrow q$ decreases.

In Fig. 2 we present two trajectories of q describing the evolution of the system from the initial equilibrium state when two different forcings, one subcritical and one supercritical, are applied. The oscillations in the value of q, which are slowly decaying in the subcritical case, are caused by the action of the heat and salinity advection feedbacks, which act on time scales of the order of the flushing time of the oceanic boxes ($t_s \approx 250$ yr). In the case of the supercritical perturbation, the oscillations of the value of q grow in size until the system collapses to a southern-sinking equilibrium. The two trajectories separate ~400 yr after the beginning of the perturbation, when the negative feedback due to the decrease of advection of warm, salty water in box 1 becomes of critical relevance. The threshold value for q is of the order of 10 Sv; more complex models have shown that weak THCs are not stable (Rahmstorf 1995; Tziperman 2000). When the THC does not change sign, it recovers to its initial unperturbed value (if $\Delta F_3 = 0$) or overshoots it ($\Delta F_3 > 0$), as can be deduced from (11); this reminds us of results obtained with much more complex models (Wiebe and Weaver 1999).

a. Hysteresis

The hysteresis describes the response of the THC against quasi-static freshwater flux perturbations, that is, with a very large t_0, in both boxes 1 and 3. In Fig. 3 we have plotted four curves, corresponding to $\Delta F_3/\Delta F_1 = [0, 0.1, 0.2, 0.3]$. The obtained graphs essentially represent the stationary states for given values of the freshwater flux. For a given curve, each point (x, y) plotted represents a stable state of the system having $q = y$ and $F_1 = x$. We scale all the graphs so that the common initial equilibrium state is the point (1, 1). For a range of F_1, which depends on the value of $\Delta F_3/\Delta F_1$, the system is bistable; that is, it possesses two distinct stable states, one having $q > 0$, the other one having $q < 0$. The history of the system determines which of the two stable states is realized. The boundaries of the domain in F_1 where the system is bistable correspond to subcritical Hopf bifurcations (Scott et al. 1999) that drive the system from the northern (southern)-sinking equilibrium to the southern (northern)-sinking equilibrium if F_1 crosses the right (left) boundary of the domain of bistability. If $\Delta F_3/\Delta F_1$ is large enough, slow increases of the freshwater fluxes do not destabilize the system, so that no bifurcations are present. It is interesting to note that if F_3 is kept constant, the value of q does not depend on the value of F_1 in the northern-sinking branch, while clearly the dependence is apparent in the south-

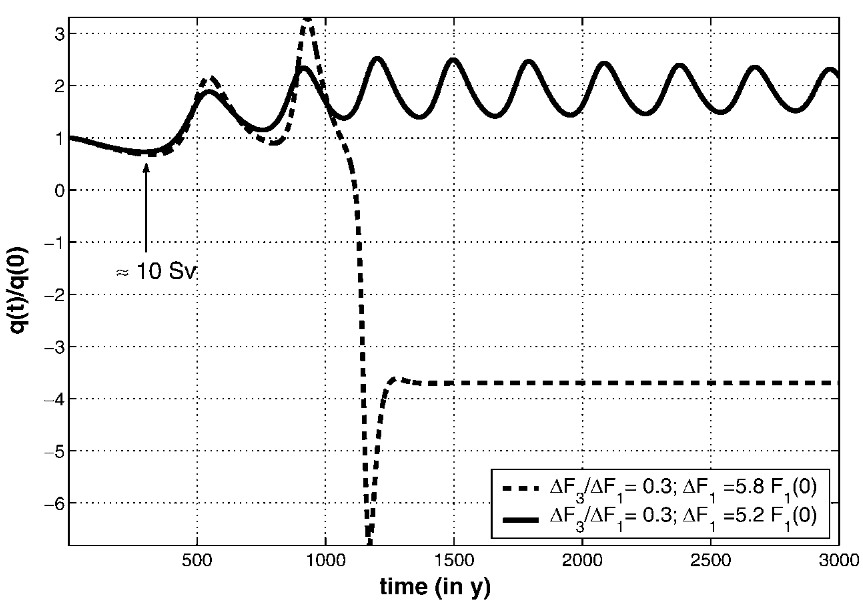

FIG. 2. Evolution of the THC strength under a super- and subcritical freshwater flux forcing.

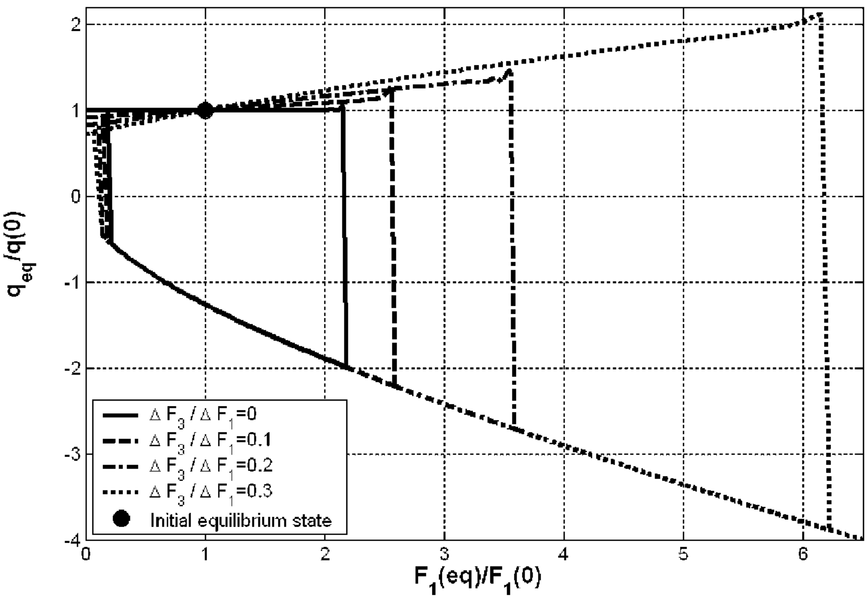

FIG. 3. Hysteresis graph of freshwater flux forcings.

ern-sinking branch. This can be explained by considering that at equilibrium, the value of q is mainly determined by the value of the freshwater flux into the upwelling box, as shown in Eq. (11). We note that since these hysteresis experiments involve simultaneous perturbations to the moisture fluxes in both hemispheres, they differ from those presented in Rahmstorf (1996), where the perturbations to the moisture fluxes occurred in one high-latitude box.

Albeit with differences in the experimental settings in some cases, the presence of a hysteresis for slow freshwater flux perturbations has been previously presented both for simple (Rahmstorf 1996; Scott et al. 1999; Titz et al. 2002b) and more complex interhemispheric models (Stocker and Wright 1991; Mikolajewicz and Maier-Reimer 1994; Rahmstorf 1995, 1996), while the inclusion of changes in the freshwater flux into both high-latitude boxes as presented here had not been previously considered.

b. *Critical forcings*

Figures 4a and 4b present, using the coordinates described by the triplets 1 and 2, respectively, the contours of the height of the manifold dividing the subcritical from the supercritical forcings. We consider logarithmic scales instead of linear scales since the former allow us to capture more clearly the qualitative behavior of the system. In Fig. 4a we observe that, as expected, the lower the value of the ratio $\Delta F_3/\Delta F_1$, the lower the total change ΔF_1 needed to obtain the reversal of the THC. For a given value of the ratio $\Delta F_3/\Delta F_1$, more rapidly increasing perturbations (larger F_1^t) are more effective in disrupting the circulation. For values of $\Delta F_3/\Delta F_1 \leq 0.4$ we see a changeover between a slow and a fast regime, geometrically described by a relatively high-gradient region dividing two portions of the manifold having little x dependence. Figure 4b shows more clearly that for large values of $\Delta F_3/\Delta F_1$ ($\geq$0.5), the collapse of the THC occurs only for fast increases because of the presence of a threshold in the rate of increase depicted by the little t_0 dependence of the manifold. The changeover corresponds to $t_0 \approx t_s$, which essentially matches the flushing time of the oceanic boxes (Scott et al. 1999). For low values of $\Delta F_3/\Delta F_1$, the disruption of the initial THC pattern takes place even in the case of a very low F_1^t, as essentially captured in the hysteresis study of the previous subsection. This occurs because, as shown in the analysis performed by Scott et al. (1999), there is a critical ratio F_3/F_1 below which the northern-sinking equilibrium becomes unstable. Such a ratio can be reached after a long enough t_0 if the slow perturbation is characterized by a low enough value of $\Delta F_3/\Delta F_1$.

c. *Relevance of the mixed boundary conditions*

In order to explore the relevance of the mixed boundary conditions (BCs) for the stability of the THC pattern against freshwater flux perturbations (Tziperman et al. 1994; Nakamura et al. 1994), we have also considered a second version of the model where we replace the expression of the heat flux in the box i defined as $\lambda(\tau_i - T_i)$ with a free parameter H_i, thus realizing a flux BC model. Choosing the value of the initial H_i equal to the value of $\lambda(\tau_i - T_i)$ of the previous version of the model at equilibrium, we obtain the same equilibrium solution. We then perturb this equilibrium

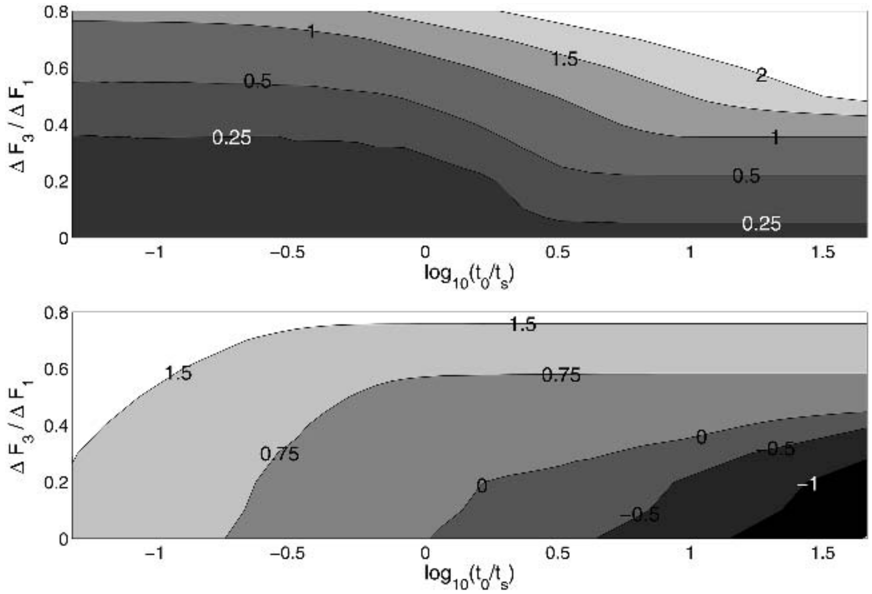

FIG. 4. Stability of the THC against freshwater flux perturbations. (a) Critical values of the total increase of the freshwater flux; contours of $\log_{10}[\Delta F_1/F_1(0)]$. (b) Critical values of the rate of increase of the freshwater flux; contours of $\log_{10}[F_1' t_0/F_1(0)]$, where t_0 is the total length of forcing.

solution with freshwater flux forcing as in the previous case. Since the heat flux H_i does not depend on T_i, the positive feedback (condition 1b) described at the beginning of the section is off; therefore we expect this version of the system to be more stable than the previous one. Repeating the analysis of stability to freshwater flux perturbations for the flux BC model, it is indeed confirmed that temperature-restoring conditions decrease the stability of the model, in agreement with the results of Tziperman et al. (1994), Nakamura et al. (1994), and Rahmstorf (2000).

Comparing the hysteresis graphs obtained for the mixed and flux BC systems presented in Fig. 5, it is apparent that the mixed BC model is less stable with

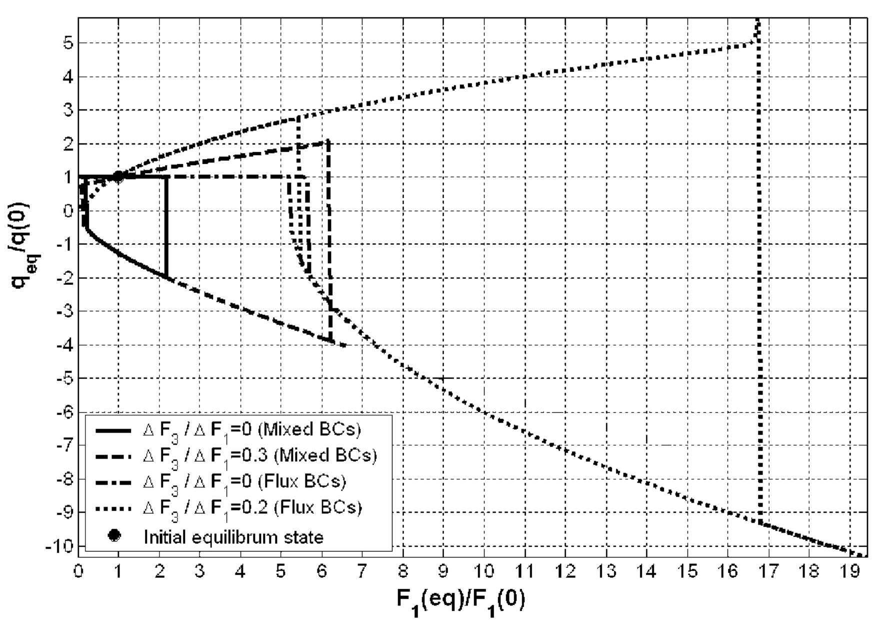

FIG. 5. Comparison between the stability of the mixed and flux BC uncoupled model to increases of the freshwater fluxes; hysteresis graphs.

respect to freshwater flux perturbations. Excluding the temperature feedback mechanism, we do not obtain collapses of the THC for quasi-static perturbations for $\Delta F_3/\Delta F_1 = 0.3$, while for each of the cases $\Delta F_3/\Delta F_1 = [0, 0.1, 0.2]$ the bifurcation point is much farther than in the mixed BC model from the point (1, 1), which describes the unperturbed system.

Figure 6 shows the ratio between the values of the critical values of ΔF_1 presented in Fig. 4 and the corresponding values obtained with the flux BC model. The average value of this ratio over all the domain is ≈ 0.3. Observing Fig. 6 for $\Delta F_3/\Delta F_1 = [0.3, 0.4]$, we see that the ratio goes to zero with increasing t_0 because in this subdomain for slow forcings we have destabilization for finite values of ΔF_1 only in the mixed BC case. This occurs because for these values the threshold in the rate of increase is present only in the flux BC model; therefore in this model for large t_0, the quantity ΔF_1 diverges, while in the mixed BC model it is finite.

6. Thermal forcing

We explore the stability of the northern-sinking equilibrium against perturbations to the initial heat fluxes expressed as changes in the target temperatures in the three boxes. We do not impose a constraint fixing the average target temperature τ_M, so that in this case not two but three parameters τ_1, τ_2, τ_3 can be changed. It is possible to recast the system of Eqs. (1)–(6) (Scott et al. 1999) so that the heat forcing is described completely by the changes in $\tau_N = \tau_1 - \tau_2$ and in $\tau_S = \tau_3 - \tau_2$. Therefore, we set $\tau_2 = \tau_2(0)$ at all times, so that $\tau_1^t = \tau_N^t$ and $\tau_3^t = \tau_S^t$. We apply the following forcings:

$$\tau_i(t) = \begin{cases} \tau_i(0) + \tau_i^t t, & 0 \le t \le t_0 \quad i = 1, 3 \\ \tau_i(0) + \tau_i^t t_0 & t > t_0 \quad i = 1, 3 \end{cases}. \tag{17}$$

We characterize each forcing experiment, which leads to a final state that is characterized by a northern- or a southern-sinking equilibrium, using the parameter set 1 defined as $[t_0, \Delta\tau_3/\Delta\tau_1, \Delta\tau_1]$ and the parameter set 2 defined as $[t_0, \Delta\tau_3/\Delta\tau_1, \tau_1^t]$; consistently with the previous case, we define $\Delta\tau_i \equiv \tau_i^t t_0$, that is, the total change of the target temperature of the box i. In order to imitate global warming scenarios where polar amplification occurs, we consider $\Delta\tau_i \ge 0$, $i = 1, 3$. We observe that if $\Delta\tau_3 > \Delta\tau_1$, no destabilization of the THC can occur because such a forcing reinforces the THC. The specified forcings trigger the previously described feedbacks of the system so that: τ_1's increase is larger than τ_3's $\Rightarrow$ H_1's increase is larger than H_3's $\Rightarrow$ T_1 increases more than $T_3 \Rightarrow q$ decreases.

We present in Fig. 7 two trajectories of q starting from the initial equilibrium state, characterized by slightly different choices of the forcing applied to τ_1 and τ_3. One of the forcings is supercritical, the other one is subcritical. The two trajectories are barely distinguishable up to the end of the perturbation, because the forcing dominates the internal feedbacks of the system thanks to the large value of λ. The oscillations in the value of q at the end of the perturbation are caused by the internal feedbacks of the system and have a period comparable with the time scale of the flushing of the oceanic boxes. In general, the mean flow dampens the increase in T_1 and tends to keep the system in the northern-sinking state, although the decreasing q low-

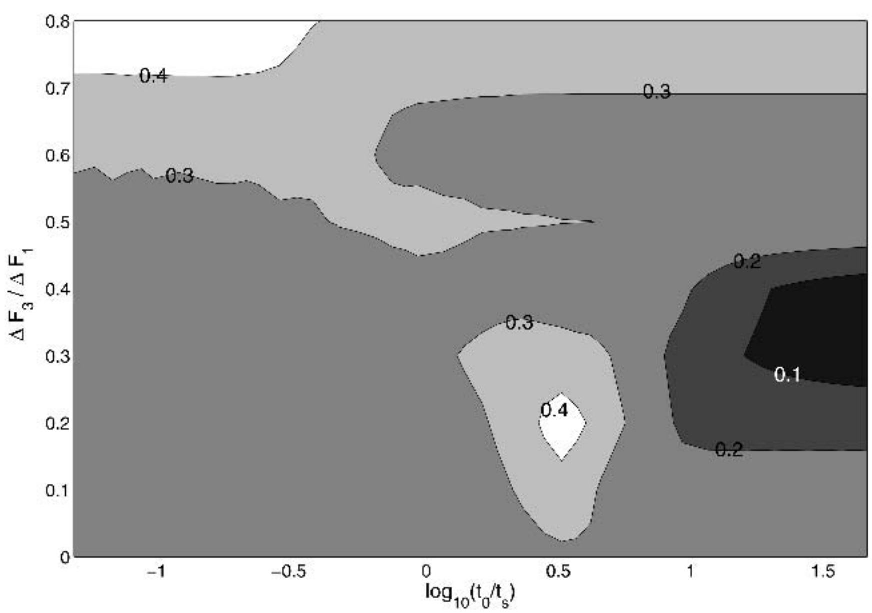

FIG. 6. Comparison between the stability of the mixed and flux BC uncoupled model to increases of the freshwater fluxes; contours of $\Delta F_1^{\mathrm{mix}}/\Delta F_1^{\mathrm{flux}}$.

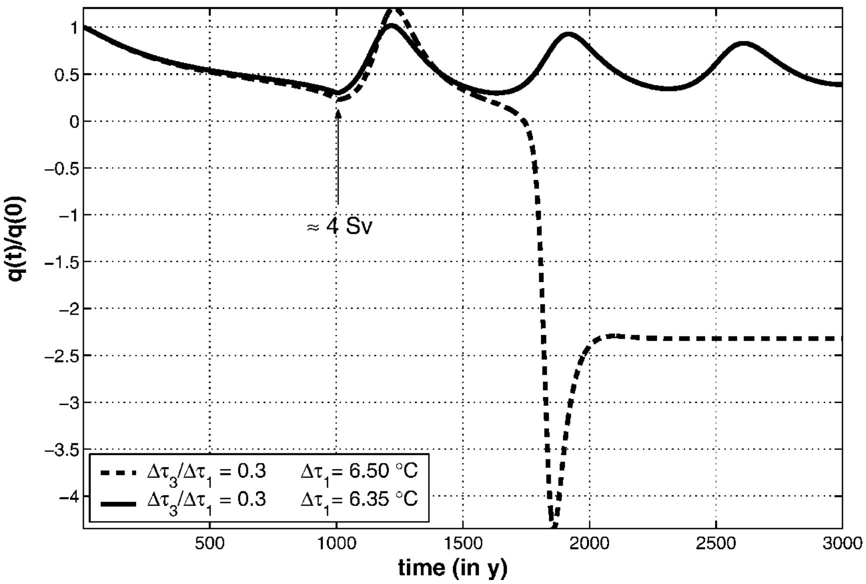

FIG. 7. Evolution of the THC strength under a super- and subcritical forcing to target temperatures.

ers the amount of warm, salty tropical water injected into box 1. When q gets too small (in this case about 4 Sv) the mean flow feedback becomes negligible and the system eventually collapses.

a. Hysteresis

We present in Fig. 8 only the graphs relative to $\Delta\tau_3/\Delta\tau_1 = [0, 0.2, 0.5, 0.8]$ for the sake of simplicity; the other curves that can be obtained for other values of $0 \leq \Delta\tau_3/\Delta\tau_1 < 1$ are not qualitatively different. The initial state is the point (0, 1). The subcritical Hopf bifurcation points are (~5, ~0.7), (~6, ~0.7) (~9, ~0.7), and (~18, ~0.6), respectively. This means that the bifurcation occurs when the THC has already declined down to ~10 Sv. The bistable region is extremely large in all cases, because the x of the bifurcation points in the lower part of the circuits are below −50°C; they become farther and farther from the initial equilibrium

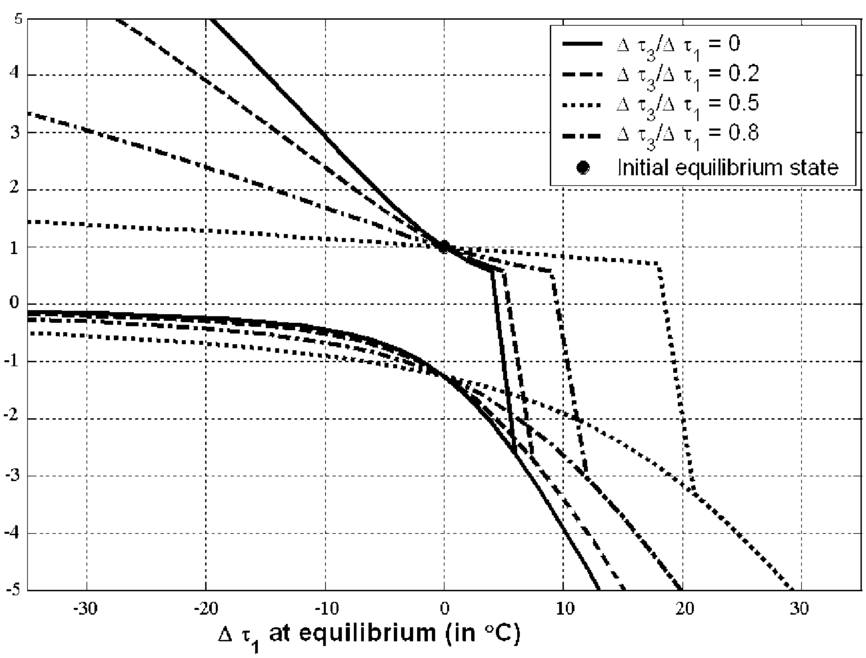

FIG. 8. Hysteresis graph of the target temperature forcings.

value as $\Delta\tau_3/\Delta\tau_1$ increases. This means that once the circulation has reversed, the newly established pattern is extremely stable and can hardly be changed again. To our knowledge, this is the first time such an analysis of hysteresis graphs for this class of forcing is presented in the literature.

b. Critical forcings

Figure 9 presents the contours of the height of the manifold of the critical forcings. We can observe that for all the values of $\Delta\tau_3/\Delta\tau_1$ there is a high-gradient region dividing two flat regions describing the fast and slow regimes. In the fast region, a smaller total increase $\Delta\tau_1$ is required to achieve the reversal of the THC. For $\Delta\tau_3/\Delta\tau_1 = 0.9$, which represents the case of highly symmetrical forcings, the critical $\Delta\tau_1$ essentially does not depend on t_0. We emphasize that the time scale of such a changeover is larger than $t_s \approx 250$ yr, the main reason being that because of the large value of λ, which is about one order of magnitude larger than the initial q, the forcing is stronger than any negative feedback even if it is relatively slow. For the same reason, in the case of perturbations to the target temperatures, we do not observe the presence of thresholds for the rate of increase of the perturbations as seen for freshwater forcings. If the increase in τ_1 is slow enough, the negative feedbacks make the destabilization more difficult, but do not prevent it.

7. Conclusions

In this paper we have accomplished a thorough analysis of the stability of the THC as described by the Rooth model. We simulate global warming conditions by applying to the equilibrium state perturbations to the moisture and heat fluxes into the three boxes. In particular, the relevance of the role of changes in the freshwater flux into the Southern Ocean is determined, along the lines of Scott et al. (1999) for a similar box model and Wang et al. (1999a) for an OGCM.

We have presented the hysteresis graphs for freshwater flux perturbations that reproduce and extend the results previously given in the literature, because the effect of freshwater flux increases in the Southern Ocean are included, while for the first time hysteresis experiments for heat flux forcings, obtained by perturbation in the target temperature, are presented.

High rates of increase in the moisture flux into the northern high-latitude box lead to a THC breakdown at smaller final increases than low rates, while the presence of moisture flux increases into the southern high-latitude box strongly inhibit the breakdown. Similarly, fast heat flux increases in the Northern Hemisphere destabilize the system more effectively than slow ones, and again the enhancement of the heat fluxes in the Southern Hemisphere tend to drive the system toward stability. In all cases analyzed, slow forcings, if asymmetric enough, lead to the reversal of the THC.

In our model, the spatial pattern of forcing is critical in determining the response of the system: quasi-symmetric forcings, which seem more reasonable under realistic global warming conditions, are not very effective in destabilizing the THC. The simplicity of the feedbacks characterizing our model, and in particular the absence of stratification effects, may be responsible for this behavior.

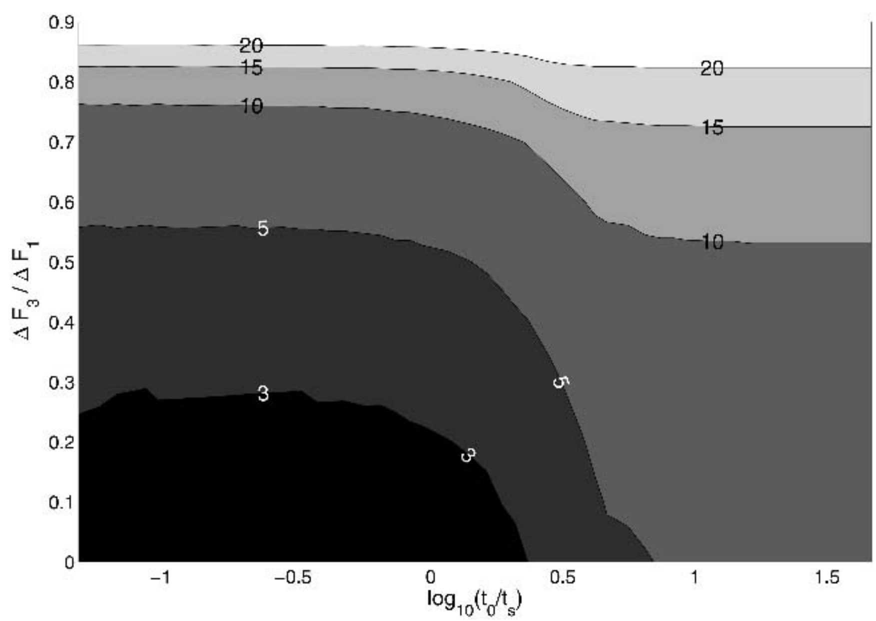

FIG. 9. Stability of the THC against target temperature perturbations. Critical values of the total increase of the target temperature τ_1; contours of $\Delta\tau_1$ in °C.

We have shown that the restoring boundary conditions for temperature decrease the stability of the system because they introduce an additional positive feedback that tends to destabilize the system. Flux boundary conditions provide the system with much greater stability with respect to perturbations in the freshwater fluxes at all time scales.

Our study is based on a conceptual model; therefore it aims at obtaining a qualitative picture of the properties of the THC more than providing realistic results. Nevertheless, future increases in computer power availability could allow the adoption of this methodology for performing extensive parametric studies of the THC stability properties and evolution with more complex and realistic models in both paleoclimatic and global warming perspectives.

We conclude that the study of the THC response to perturbations performed with uncoupled models, of any level of complexity, should take into much greater account the spatial pattern of the freshwater and/or heat flux forcing. The THC is a highly nonlinear, nonsymmetric system, and we suggest that the effect of changing the rate of increase of the forcings should be explored in greater detail to provide a bridge between the instantaneous and quasi-static changes. The THC dynamics encompass very different time scales, which can be explored only if the timing of the forcings is varied.

We conclude by pointing out some possible improvements to the present model and possible extensions of this work. We have put emphasis on how the oceanic advection can reverse the symmetry of the pattern of the circulation in the context of a rigorously symmetric geometry. The system is likely to be very sensitive to changes in the volumes of the boxes and especially to the introduction of asymmetries between the two high-latitude boxes, which would also make the system more realistic since the southern mid- to high-latitude portion of the Atlantic is considerably larger than the northern mid- to high-latitude portion (Zickfeld et al. 2004). Asymmetries in the oceanic fractional areas would induce asymmetries in the values of λ_i, thus causing the presence of different restoring times for the various boxes.

The introduction of a noise component in the tendency equation would increase the model's realism and make the model conceptually more satisfying since it would introduce a parameterization of the high-frequency variability. Recent studies have undertaken this strategy and shown that close to instability threshold the evolution of the THC has a very limited predictability (Wang et al. 1999a; Knutti and Stocker 2002), and that stochastic resonance could be responsible for glacial/interglacial climate shifts (Ganopolski and Rahmstorf 2002). It would be interesting to analyze how the intensity and color of the noise would influence the results obtained and the conclusions drawn in this work.

Finally, the model could be improved so that it would be able to represent the main features of the whole conveyor belt by adding other boxes, descriptive of other oceanic basins, and carefully setting the connections between the boxes, along the lines of some examples presented by Weaver and Hughes (1992). Moreover, the consideration of a deep ocean box would surely increase the model performance in terms of representation of the short time scales.

In an extension of this paper (Lucarini and Stone 2005), we perform a similarly structured stability analysis of the THC under global warming conditions using a more complex version of the Rooth model, where an explicit coupling between the oceanic and the atmospheric part is introduced and the main atmospheric physical processes responsible for freshwater and heat fluxes are formulated separately.

Acknowledgments. The authors are grateful to J. Scott for interesting discussions and useful suggestions. One author (V.L.) wishes to thank R. Stouffer for having proposed improvements to an earlier version of the manuscript, and T. Stocker and E. Tziperman for having suggested a number of relevant references. This research was funded in part by the U.S. Department of Energy's (DOE's) Climate Change and Prediction Program and in part by the MIT Joint Program on the Science and Policy of Global Change (JPSPGC). Financial support does not constitute an endorsement by DOE or JPSPGC of the views expressed in this article.

REFERENCES

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

Boyle, E. A., and L. Keigwin, 1987: North Atlantic thermohaline circulation during the past 20 000 years linked to high-latitude surface temperature. *Nature,* **330,** 35–40.

Bretherton, F. P., 1982: Ocean climate modeling. *Progress in Oceanography,* Vol. 11, Pergamon, 93–129.

Broecker, W. S., 1997: Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO_2 upset the current balance? *Science,* **278,** 1582–1588.

——, D. M. Peteet, and D. Rind, 1985: Does the ocean–atmosphere system have more than one stable mode of operation? *Nature,* **315,** 21–25.

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

Bugnion, V., and C. Hill, 2003: The equilibration of an adjoint model on climatological time scales. Tech. Rep. 72, Center for Global Change Science, MIT, 156 pp.

Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 526–582.

Ganopolski, A., and S. Rahmstorf, 2002: Abrupt glacial climate changes due to stochastic resonance. *Phys. Rev. Lett.,* **88,** 038 501, doi:10.1103/PhysRevLett.88.038501.

Huang, R. X., 1999: Mixing and energetics of the oceanic thermohaline cirulation. *J. Phys. Oceanogr.,* **29,** 727–746.

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

Keigwin, L. D., W. B. Curry, S. J. Lehman, and S. Johnsen, 1994: The role of the deep ocean in North Atlantic climate change between 70 and 130 kyr ago. *Nature,* **371,** 323–326.

Klinger, B. A., and J. Marotzke, 1999: Behavior of double-hemisphere thermohaline flows in a single basin. *J. Phys. Oceanogr.,* **29,** 382–400.

Knutti, R., and T. F. Stocker, 2002: Limited predictability of the future thermohaline circulation close to an instability threshold. *J. Climate,* **15,** 179–186.

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

Latif, M., E. Roeckner, U. Mikolajewicz, and R. Voss, 2000: Tropical stabilization of the thermohaline circulation in a greenhouse warming simulation. *J. Climate,* **13,** 1809–1813.

Lucarini, V., and P. H. Stone, 2005: Thermohaline circulation stability: A box model study. Part II: Coupled model. *J. Climate,* **18,** 514–529.

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

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

——, and ——, 1993: Century-scale effects of increased atmospheric CO_2 on the ocean–atmosphere system. *Nature,* **364,** 215–218.

——, and ——, 1999a: Are two modes of thermohaline circulation stable? *Tellus,* **51A,** 400–411.

——, and ——, 1999b: The role of thermohaline circulation in climate. *Tellus,* **51A–B,** 91–109.

Marotzke, J., 1996: Analysis of thermohaline feedbacks. *Decadal Climate Variability: Dynamics and Predicatibility,* Springer, 333–378.

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

——, and P. H. Stone, 1995: Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *J. Phys. Oceanogr.,* **25,** 1350–1360.

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

Munk, W., and C. Wunsch, 1998: Abyssal recipes. II. Energetics of the tides and wind. *Deep-Sea Res.,* **45,** 1976–2009.

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

Peixoto, A., and B. Oort, 1992: *Physics of Climate.* American Institute of Physics, 520 pp.

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

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

——, 1997: Risk of sea-change in the Atlantic. *Nature,* **388,** 825–826.

——, 1999a: Rapid oscillation of the thermohaline ocean circulation. *Reconstructing Ocean History: A Window into the Future,* Kluwer Academic, 309–332.

——, 1999b: Shifting seas in the greenhouse? *Nature,* **399,** 523–524.

——, 2000: The thermohaline ocean circulation—A system with dangerous thresholds? *Climatic Change,* **46,** 247–256.

——, 2002: Ocean circulation and climate during the past 120 000 years. *Nature,* **419,** 207–214.

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

——, and M. H. England, 1997: Influence of Southern Hemisphere winds on North Atlantic deep water flow. *J. Phys. Oceanogr.,* **27,** 2040–2054.

Roemmich, D. H., and C. Wunsch, 1985: Two transatlantic sections: Meridional circulation and heat flux in the subtropical North Atlantic Ocean. *Deep-Sea Res.,* **32,** 619–664.

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

Sandström, J. W., 1908: Dynamische Versuche mit Meerwasser. *Ann. Hydrogr. Maritimen Meteor.*, **36,** 6–23.

Scott, J. R., J. Marotzke, and P. H. Stone, 1999: Interhemispheric thermohaline circulation in a coupled box model. *J. Phys. Oceanogr.,* **29,** 351–365.

Seager, R., D. S. Battistini, J. Yin, N. Gordon, N. Naik, A. C. Clement, and M. A. Cane, 2002: Is the gulf stream responsible for Europe's mild winters? *Quart. J. Roy. Meteor. Soc.,* **128,** 2563–2586.

Stocker, T. F., 2001: The role of simple models in understanding climate change. *Continuum Mechanics and Applications in Geophysics and the Environment,* Springer, 337–367.

——, and D. G. Wright, 1991: Rapid transitions of the ocean's deep circulation induced by changes in the surface water fluxes. *Nature,* **351,** 729–732.

——, and A. Schmittner, 1997: Influence of CO_2 emission rates on the stability of the thermohaline circulation. *Nature,* **388,** 862–864.

——, R. Knutti, and G.-K. Plattner, 2001: The future of the thermohaline circulation—A perspective. *The Oceans and Rapid Climate Change: Past, Present and Future,* Geophys. Monogr., Vol. 126, Amer. Geophys. Union, 277–293.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus,* **13,** 224–227.

Stone, P. H., and Y. P. Krasovskiy, 1999: Stability of the interhemispheric thermohaline circulation in a coupled box model. *Dyn. Atmos. Oceans,* **29,** 415–435.

Stouffer, R. J., S. Manabe, and K. Bryan, 1991: Climatic response to a gradual increase of atmospheric carbon dioxide. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations,* Elsevier, 545–557.

Titz, S., T. Kuhlbrodt, and U. Feudel, 2002a: Homoclinic bifurcation in an ocean circulation box model. *Int. J. Bif. Chaos,* **12,** 869–875.

——, ——, S. Rahmstorf, and U. Feudel, 2002b: On freshwater-dependent bifurcations in box models of the interhemispheric thermohaline circulation. *Tellus,* **54A,** 89–98.

Toggweiler, J. R., and B. Samuels, 1995: Effect of drake passage on the global thermohaline circulation. *Deep-Sea Res.,* **42A,** 477–500.

Tziperman, E., 2000: Proximity of the present-day thermohaline circulation to an instability threshold. *J. Phys. Oceanogr.,* **30,** 869–875.

——, and H. Gildor, 2002: The stabilization of the thermohaline circulation by the temperature–precipitation feedback. *J. Phys. Oceanogr.,* **32,** 2707–2714.

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

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

——, ——, and ——, 1999b: Thermohaline circulation. Part II: Sensitivity with interactive atmospheric transport. *J. Climate,* **12,** 83–92.

Weaver, A. J., and T. M. C. Hughes, 1992: Stability of the thermohaline circulation and its links to climate. *Trends in Physical Oceanography*, Council of Scientific Research Integration, Ed., Research Trends Series, Vol. 1, 15–70.

Wiebe, E. C., and A. J. Weaver, 1999: On the sensitivity of global warming experiments to the parametrisation of sub-grid scale ocean mixing. *Climate Dyn.,* **15,** 875–893.

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.

Zickfeld, K., T. Slawig, and S. Rahmstorf, 2004: A low-order model for the response of the Atlantic thermohaline circulation to climate change. *Ocean Dyn.,* **54,** 8–26.

Thermohaline Circulation Stability: A Box Model Study. Part II: Coupled Atmosphere–Ocean Model

VALERIO LUCARINI*

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

PETER H. STONE

Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 23 June 2003, in final form 29 July 2004)

ABSTRACT

A thorough analysis of the stability of a coupled version of an interhemispheric three-box model of thermohaline circulation (THC) is presented. This study follows a similarly structured analysis of an uncoupled version of the same model presented in Part I of this paper. The model consists of a northern high-latitude box, a tropical box, and a southern high-latitude box, which can be thought of as corresponding to the northern, tropical, and southern Atlantic Ocean, respectively. This paper examines how the strength of THC changes when the system undergoes forcings representing global warming conditions.

Since a coupled model is used, a direct representation of the radiative forcing is possible because the main atmospheric physical processes responsible for freshwater and heat fluxes are formulated separately. Each perturbation to the initial equilibrium is characterized by the total radiative forcing realized, by the rate of increase, and by the north–south asymmetry. Although only weakly asymmetric or symmetric radiative forcings are representative of physically reasonable conditions, general asymmetric forcings are considered in order to get a more complete picture of the mathematical properties of the system. The choice of suitably defined metrics makes it possible to determine the boundary dividing the set of radiative forcing scenarios that lead the system to equilibria characterized by a THC pattern similar to the present one, from those that drive the system to equilibria where the THC is reversed. This paper also considers different choices for the atmospheric transport parameterizations and for the ratio between the high-latitude and tropical radiative forcing. It is generally found that fast forcings are more effective than slow forcings in disrupting the present THC pattern, forcings that are stronger in the northern box are also more effective in destabilizing the system, and very slow forcings do not destabilize the system whatever their asymmetry, unless the radiative forcings are very asymmetric and the atmospheric transport is a relatively weak function of the meridional temperature gradient. In this latter case some relevant hysteresis graphs of the system are presented. The changes in the strength of the THC are primarily forced by changes in the latent heat transport in the hemisphere because of its sensitivity to temperature, which arises from the Clausius–Clapeyron relation.

1. Introduction

This analysis follows the study reported in the first part of this paper (Lucarini and Stone 2005, hereafter Part I), where we dealt with an uncoupled three-box oceanic model. The reader should refer to the introduction of Part I for a brief presentation of the concept of thermohaline circulation (THC) and for a discussion of the relevant issues regarding the THC's fate in the context of global warming as well in a paleoclimatic perspective.

The model analyzed here is characterized by the presence of explicit coupling between the ocean and the atmosphere. The atmospheric freshwater and heat fluxes are expressed as functions of the oceanic temperatures. We emphasize that the analytic formulation of the freshwater flux includes the Clausius–Clapeyron effect, while the total atmospheric heat flux is given by the sum of three functionally distinct contributions, representing the latent, sensible, and radiative heat fluxes. The more detailed mathematical description of the heat transfer processes allows us to consider a conceptually more precise picture of global warming scenarios, where the external radiative forcing acts on the radiative heat flux term alone. Previous studies have

* Current affiliation: Dipartimento di Matematica e Informatica, University of Camerino, Camerino Italy.

Corresponding author address: Dr. Valerio Lucarini, Dept. of Mathematics and Computer Science, University of Camerino, via Madonna delle Carceri, 62032 Camerino (MC), Italy.
E-mail: lucarini@alum.mit.edu

shown that the interaction with the atmosphere often has a strong effect of the stability properties of the THC (Nakamura et al. 1994; Scott et al. 1999; Wang et al. 1999b).

In this paper we perform a parametric study of the relevance of both spatial and temporal patterns of the radiative forcing and characterize the response of the system by determining the thresholds beyond which we have destabilization of the present mode of the THC and transition to a reversed circulation. We consider an extremely simplified climate model. On one side, this allows a very thorough exploration of the parameters' space. On the other side, this limits the scope of the paper to providing qualitative information that should be considered as conceptual and methodological suggestions for more detailed analyses to be performed using more complex models.

Our paper is organized as follows: in section 2 we provide a brief description and the general mathematical formulation of the dynamics of the three-box model used in this study and we describe the parameterization of the coupling between the atmosphere and the ocean. In section 3 we describe the forcings we apply. In section 4 we analyze some relevant model runs. In section 5 we present a study of some relevant hysteresis graphs of the system. In section 6 we treat the general stability properties of the system. In section 7 we perform a parametric sensitivity study. In section 8 we present our conclusions.

2. Brief model description

The three-box model consists of a northern high-latitude box (box 1), a tropical box (box 2), and a southern high-latitude box (box 3). The volume of the two high-latitude boxes is the same and is $1/V$ times the volume of the tropical box. We choose $V = 2$, so that box 1, box 2, and box 3 can be thought of as describing the portions of an ocean like the Atlantic north of 30°N, between 30°N and 30°S, and south of 30°S, respectively. The two high-latitude boxes are connected by a deep current passage of negligible mass. The boxes are well mixed, and T_i and S_i are the temperature and salinity of the box i, respectively, while $\tilde{F}_i$ and $\tilde{H}_i$ are the net freshwater and atmospheric heat fluxes into box i, respectively. The box i is subjected to the oceanic advection of heat and salt from the upstream box through the THC, whose strength is $\tilde{q}$. The atmosphere and the land have a negligible heat capacity and water content compared to the ocean. The structure of the model we consider in this paper is the same as the model analyzed in Part I. We suggest that the reader refer to the description of the model, the notation, the discussion of the tendency equations, and their nondimensionalization in Part I. Note that the transport quantities without the tilde are suitably nondimensionalized counterparts of the physical variables.

a. *Parameterization of the coupling between the atmosphere and the ocean*

The model we analyze in this work is formulated in a different way with respect to Part I since it incorporates an explicit coupling between the ocean and the atmosphere. The atmospheric fluxes of heat and freshwater are parameterized as functions of the box temperatures. We choose simple but physically plausible functional forms that are based on the large-scale processes governing the transfer of heat and freshwater through the atmosphere. We want to capture the dependence of atmospheric transport from the Tropics to the high latitudes on the temperature gradient, considering that baroclinic eddies contribute to most of the meridional transport around 30°N (Peixoto and Oort 1992), the dependence of the outgoing longwave radiation on the temperature, and the dependence of the moisture content of the atmosphere on the temperature. The last property differentiates the parameterization we choose from the otherwise closely similar Scott et al. (1999) model's choices.

b. *Freshwater fluxes*

The nondimensionalized net freshwater fluxes F_i are parameterized following Stone and Miller (1980) and Stone and Yao (1990):

$$F_i = \frac{C_i}{\left(\frac{T_2 + T_i}{2}\right)^3} \exp\left[-\frac{L_v}{R_v \frac{T_2 + T_i}{2}}\right](T_2 - T_i)^n, \quad i = 1, 3, \tag{1}$$

$$F_2 = -\frac{1}{V}(F_1 + F_3), \tag{2}$$

where L_v is the unit mass latent heat of vaporization of water (taken as constant), R_v is the gas constant, and C_1 and C_3 are the coefficients we have to calibrate. The exponential functions are derived from the Clausius–Clapeyron law, while the value of the exponent n determines the sensitivity of the process of baroclinic eddy transport on the meridional temperature gradient. As in Part I, F_2 is obtained by imposing the conservation of the salinity.

We note that the Clausius–Clapeyron equations had been included in the description of atmosphere–ocean coupling in earlier studies performed on box models, but those dealt with hemispheric and not interhemispheric models (Nakamura et al. 1994; Tziperman and Gildor 2002).

c. *Heat fluxes*

The surface heat fluxes $\tilde{H}_i$ are decomposed in three components describing physically different phenomena,

$$\tilde{H}_i = \widetilde{\mathrm{LH}}_i + \widetilde{\mathrm{SH}}_i + \widetilde{\mathrm{RH}}_i, \tag{3}$$

where $\widetilde{\mathrm{LH}}_i$ and $\widetilde{\mathrm{SH}}_i$ are the convergence of the atmospheric flux of latent and sensible heat in the box i, respectively, while $\widetilde{\mathrm{RH}}_i$ describes the radiative balance between incoming solar radiation and outgoing longwave radiation. The convergence of atmospheric transports must globally sum up to zero at any time, since the atmosphere is closed.

1) Latent heat flux

The oceanic box $i = 1, 3$ receives a fraction $1/\gamma_i$ of the total net moisture transported by the atmosphere from the Tropics to high latitudes in the Northern ($i = 1$) and Southern Hemisphere ($i = 3$), respectively; this includes the fraction that directly precipitates over the oceanic boxes $i = 1, 3$ and the fraction that precipitates over land and runs off to the oceanic box $i = 1, 3$.

The fractional catchment area $1/\gamma_i$ can range in our system from 1 (all of the atmospheric moisture exported from the Tropics to the high latitudes ends up in the box $i = 1, 3$, respectively) to 1/6 (the box $i = 1, 3$ receives only the moisture transported from the Tropics that precipitates on the ocean surface). The remaining fraction $(1 - 1/\gamma_i)$ of the total atmospheric moisture exported from the Tropics returns to oceanic box 2 by river runoff or underground flow. This latter fraction does not affect the moisture budget of the oceanic boxes $i = 1, 3$ but does affect their heat budget since the process of condensation occurs over the high-latitude regions $i = 1, 3$, so that latent heat is released to the atmosphere and is immediately transferred to the oceanic box $i = 1, 3$. We then ascertain that freshwater and the latent heat fluxes are related as follows:

$$\widetilde{\mathrm{LH}}_i = \gamma_i L_v \tilde{F}_i, \quad i = 1, 3, \tag{4}$$

$$\widetilde{\mathrm{LH}}_2 = -\widetilde{\mathrm{LH}}_1 - \widetilde{\mathrm{LH}}_3, \tag{5}$$

where the total net balance of the latent heat fluxes is zero. When we apply the nondimensionalization procedure as in Part I, we find that the following relation holds between LH_i and F_i:

$$LH_i = \gamma_i \frac{L_v}{c_p S_0 \rho_0} F_i, \quad i = 1, 3. \tag{6}$$

We observe that in density units the following relation holds:

$$\left(\frac{LH_i}{F_i}\right)_\rho = \frac{\alpha \gamma_i L_v}{\beta c_p S_0} \approx 6, \quad i = 1, 3. \tag{7}$$

We underline that this value of the ratio depends on our assumed values for γ_i, α, and β. If we consider a more realistic equation of state for the density, the relative importance of freshwater and heat fluxes is expected to change in quantitative—but not qualitative—terms with respect to what was presented in Eq. (7).

2) Sensible heat flux

We consider that the baroclinic eddies are the main mechanism responsible for the meridional transport of sensible heat; therefore we assume that the sensible heat flux convergence $\widetilde{SH}_i$ is, coherent with our picture of the latent heat transport, proportional to the nth power of meridional temperature gradient (Stone and Miller 1980; Stone and Yao 1990). After the suitable nondimensionalization, we obtain the following expression for SH_i:

$$SH_i = D_i(T_2 - T_i)^n, \quad i = 1, 3,$$
$$SH_2 = -\frac{1}{V}(SH_1 + SH_2), \tag{8}$$

where SH_2 is such that the total sensible heat flux convergence globally sums up to zero.

3) Radiative heat flux

The radiative heat flux $\widetilde{RH}_i$ is modeled, as usual, as a Newtonian relaxation process (Wang and Stone 1980; Marotzke and Stone 1995; Marotzke 1996). After the correct nondimensionalization, we obtain the following expression for RH_i:

$$RH_i = A_i - B_i T_i = B_i\left(\frac{A_i}{B_i} - T_i\right) = B_i(\vartheta_i - T_i),$$
$$i = 1, 2, 3. \tag{9}$$

With respect to the box i, A_i describes the net radiative budget if $T_i = 0°\mathrm{C}$; B_i is an empirical coefficient, which, if albedo is fixed as in this model, is a measure of the sensitivity of the thermal emissions to space to surface temperature, including the water vapor and cloud feedbacks; and $\vartheta_i = A_i/B_i$ is the radiative equilibrium temperature.

4) Choice of the constants

In order to obtain for the coupled model presented here, an initial equilibrium identical to that of the uncoupled model, which is summarized in the data reported in Table 2 of Part I, we need to carefully choose the constants in the atmospheric parameterization. We first set the radiative heat flux parameters by adopting the parameterization proposed by Marotzke (1996). We set for our model $B_1 = B_2 = B_3 = B = 5.1 \times 10^{-10}$ s^{-1}. We can then obtain the following restoring equation for the global average temperature T_M:

$$\dot{T}_M = B(\vartheta_M - T_M), \tag{10}$$

where the parameter B introduces a time scale of ~60 yr for the global radiative processes (Marotzke 1996; Scott et al. 1999). The ϑ_i values are chosen following the parameterization presented in Marotzke (1996) and are such that the average global radiative temperature $\vartheta_M \equiv (\vartheta_1 + V\vartheta_2 + \vartheta_3)/(2 + V) = 15°\mathrm{C}$, which is also the value of the global average temperature at equilibrium, as in the uncoupled model in Part I. In physical terms, B corresponds to the property of the system that a global radiative forcing of 1 Wm^{-2} (which results in an effective radiative forcing of 6 Wm^{-2} in the oceanic surface fraction since we assume that land and atmosphere have no heat capacity) causes an increase of the average temperature of the system T_M of ~0.6°C when equilibrium is reestablished; this property can be summarized by introducing a climate temperature/radiation elasticity parameter $\kappa_M \approx 0.6°\ \mathrm{CW}^{-1}\ \mathrm{m}^2$. Therefore, considering that it is estimated that in the real Earth system the doubling of CO_2 causes an average radiative forcing of ~4 W m^{-2}, we can loosely interpret the parameter κ_M as indicating a model climate sensitivity of ≈2.5°C.

Substituting in Eqs. (1) and (2) the T_i of the equilibrium solution of the uncoupled model, we can derive the C_i such that we obtain $F_1 = 13.5 \times 10^{-11}$ psu s^{-1} and $F_3 = 9 \times 10^{-11}$ psu s^{-1}. The latent heat fluxes are then obtained using Eq. (6), while the coefficients D_i for the sensible heat fluxes in Eq. (8) are derived by requiring that the total heat flux H_i in (3) of the coupled and the uncoupled model match at the equilibrium solution. The relative magnitude of the latent and sensible heat fluxes LH_i and SH_i at equilibrium depend on the choice of γ_i (Marotzke 1996). Estimates for γ_i for the Atlantic range from 1.5 to 3 (Marotzke 1996). We set $\gamma_1 = \gamma_3$ in order to keep the geometry of the problem entirely symmetric, and choose $\gamma_1 = \gamma_3 = 2$.

The parameter n determines the efficiency of the atmospheric transports in terms of its sensitivity to the meridional temperature gradient. We consider for n the values [1, 3, 5], which include the domain (2–5) proposed in Held (1978), but also include simple diffusive representation ($n = 1$). We report in Table 1 the value of the main model's constants and in Table 2 the value of the quantities defining the initial equilibrium state. It is important to note that the initial equilibrium value of the THC strength introduces a natural time scale for the system $q_{\mathrm{eq}}^{-1} = t_s \approx 250$ yr, which is the flushing (or advection) time of the oceanic boxes. In particular, we observe that while the sensible and latent heat fluxes into box 1 are roughly the same; in the case of box 3 the sensible heat flux is almost 3 times as large as the latent heat flux. We note that in spite of the rather rough procedure of parameters' estimation and of deduction of the sensible and latent heat fluxes at equilibrium, the figures we obtain with our model are broadly in agreement with the climatological estimates (Peixoto and Oort 1992), as can be seen in Table 3.

3. Feedbacks of the system and radiative forcings applied

The oceanic component of our model contains the well-known oceanic feedbacks, the salinity advection (positive), and heat advection (negative) feedbacks (Scott et al. 1999; Part I). However, the atmospheric feedbacks and how they are coupled to the ocean are considerably more complex than in the analysis of Scott et al. (1999). In our case, we have to consider both changes in the meridional temperature gradients $T_i - T_2$ and in the hemispheric temperature averages $(T_i + T_2)/2$, whereas only the former was relevant in the analysis of Scott et al. (1999). In general, we expect that positive changes in the average temperature enhance the amount of moisture that can be retained by the atmosphere, thus increasing the meridional transports of moisture and latent heat. On the other hand, we expect that the reduction of meridional temperature gradients hinders the efficiency of the meridional atmospheric transports.

Table 1. Value of the main model parameters.

Quantity	Symbol	Value
Mass of box i = 1, 3	M	1.08×10^{20} kg
Box 2/box i =1, 3 mass ratio	V	2
Average density	ρ_0	1000 kg m^{-3}
Average salinity	S_0	35 psu
Oceanic fractional area	ϵ	⅙
Box i = 1,3 fractional water catchment area	$1/\gamma_i$	½
Specific heat per unit mass of water	c_p	$4 \times 10^3\ J\ °\mathrm{C}^{-1}\ \mathrm{kg}^{-1}$
Latent heat per unit mass of water	L_v	$2.5 \times 10^6\ \mathrm{J\ kg}^{-1}$
Gas constant	R_v	461 J $°\mathrm{C}^{-1}\ \mathrm{kg}^{-1}$
Thermal expansion coefficient	α	$1.5 \times 10^{-4}\ °\mathrm{C}^{-1}$
Haline expansion coefficient	β	$8 \times 10^{-4}\ \mathrm{psu}^{-1}$
Hydraulic constant	k	$1.5 \times 10^{-6}\ \mathrm{s}^{-1}$
Radiative equilibrium temperature—box 1	ϑ_1	−22.9°C
Radiative equilibrium temperature—box 2	ϑ_2	52.9°C
Radiative equilibrium temperature—box 3	ϑ_3	−22.9°C
Global climatic temperature/radiation elasticity*	κ_M	0.6°C $\mathrm{W}^{-1}\ \mathrm{m}^2$

* Coupled model: Value relative to the whole planetary surface; corresponds to a global climate sensitivity per CO_2 doubling of ~2.5°C.

Table 2. Value of the fundamental variables of the system at the initial equilibrium state.

Variable	Box 1	Box 2	Box 3
Temperature	2.9°C	28.4°C	0.3°C
Salinity	34.7 psu	35.6 psu	34.1 psu
Atmospheric freshwater flux	0.41 Sv	−0.68 Sv	0.27 Sv
Total surface heat flux	−1.58 PW	1.74 PW	−0.16 PW
Latent heat flux	1.84 PW	−3.06 PW	1.22 PW
Sensible heat flux	2.14 PW	−5.75 PW	3.61 PW
Radiative heat flux	−5.57 PW	10.57 PW	−5.00 PW
Oceanic heat flux	1.58 PW	−1.74 PW	0.16 PW
THC strength	15.5 Sv	15.5 Sv	15.5 Sv

TABLE 3. Comparison between the values of sensible and latent heat fluxes of the model at the initial equilibrium and the corresponding climatological estimates given by Peixoto and Oort (1992; hereafter PO92) for the net fluxes at 30°N and 30°S.

Variable	Box 1	Box 2	Box 3
Latent heat flux (model)	1.84 PW	−3.06 PW	1.22 PW
Latent heat flux (PO92)	~1.3 PW	~−2.6 PW	~1.3 PW
Sensible heat flux (model)	2.14 PW	−5.75 PW	3.61 PW
Sensible heat flux (PO92)	~1.5 PW	~−4 PW	~2.5 PW

The complexity of the coupled model makes it very cumbersome to present a detailed schematic description of the feedbacks of the system along the lines of Marotzke (1996) and Scott et al. (1999). The main aim of this work is to describe the behavior and stability properties of a strongly perturbed system. Therefore, instead of analyzing in detail the internal feedbacks of the unforced system, we will go directly to the study of the response of the system to the external forcing in order to capture the most relevant processes. We will then deduce which are the most relevant processes controlling the dynamics of the system by interpreting the sensitivity of the system response to some key parameters.

In our model, we simulate global warming–like radiative forcings by increasing the radiative equilibrium temperature ϑ_i. Since the dynamics of the model depends on both the averages and gradients of temperatures of neighboring boxes, we cannot limit ourselves to considering only changes in high-latitude boxes as in Part I. Therefore, changes in the parameter ϑ_2, which in the first approximation controls T_2, need to be considered in order to perform a complete and sensible study. We alter the driving parameters ϑ_i by using a linear increase:

$$\vartheta_i(t) = \begin{cases} \vartheta_i(0) + \vartheta_i^t t, & 0 \le t \le t_0 \\ \vartheta_i(0) + \vartheta_i^t t_0. & t > t_0 \end{cases} \quad i = 1, 2, 3. \tag{11}$$

We make this choice because a linearly increasing radiative forcing approximately corresponds in physical terms to an exponential increase of the concentration of greenhouse gases (Shine et al. 1995; Stocker and Schmittner 1997). The role of ϑ_2 is analyzed by considering three cases of tropical-to-high-latitude ratio of forcing (TRRF): TRRF = $\Delta\vartheta_2/\Delta\vartheta_1$ = [1.0, 1.5, 2.0], where we have used the definition $\Delta\vartheta_i = \vartheta_i^t t_0$. This allows for the fact that in global warming scenarios the net radiative forcing increase is larger in the Tropics than in mid- to high latitudes (Ramanathan et al. 1979). Such discrepancy is due to the stronger water vapor positive feedback in the Tropics caused by a larger amount of water vapor in the tropical atmosphere.

For a given value of TRRF, each forcing scenario can be uniquely identified by the triplet $[t_0, \Delta\vartheta_3/\Delta\vartheta_1, \Delta\vartheta_1]$, or alternatively by the triplet $[t_0, \Delta\vartheta_3/\Delta\vartheta_1, \vartheta_1^t]$. Although only weakly asymmetric or symmetric forcings are representative of physically reasonable conditions, in our study we consider general asymmetric forcings in order to get a more complete picture of the mathematical properties of the system and so derive a wider view of the qualitative stability properties of the THC system.

We find that in most cases the destabilization of the THC occurs if $\Delta\vartheta_1 \ge \Delta\vartheta_3$. We observe that the first-order effect of such a forcing is to weaken the THC, because the following relation holds: Increase in $\vartheta_2 \ge$ increase in $\vartheta_1 \ge$ increase in $\vartheta_3 \Rightarrow$ increase in $H_2 \ge$ increase in $H_1 \ge$ increase in $H_3 \Rightarrow$ increase in $T_2 \ge$ increase in $T_1 \ge$ increase in $T_3 \Rightarrow q$ decreases. We emphasize that in general a larger radiative forcing in the tropical box decreases the stability of the circulation because it causes advection of warmer water from box 2 to box 1.

4. Analysis of selected model runs

In Fig. 1 we show the time evolution of the THC strength q (we have chosen $n = 1$) for two slightly different symmetric radiative forcings lasting 500 yr; the solid line describes the subcritical and the dashed line the supercritical case.

In the subcritical case the forcing is such that the radiative equilibrium temperature increases by 7.5°C century^{-1} in both high-latitude boxes and by 11.25°C century^{-1} in the tropical box. This corresponds to changes in the radiative forcings of ~12 and 18 W m^{-2} century^{-1}, respectively, and so to a globally averaged increase of the radiative forcing of ~15 W m^{-2} century^{-1}.

In the subcritical case, the minimum value of THC strength (which is ~6 Sv; 1 Sv ≡ 10^6 m^3 s^{-1}) is reached at $t \approx 350$ yr. After that, and so still during the increase in the forcing, q oscillates with a period of ~400 yr. This means that the negative feedbacks overcome the external forcing and stabilize the system. We see that in the case of symmetric global warming–like radiative forcing, the short-term response of the system is characterized by an initial decrease of the THC strength, in agreement with the results of most of the models (Tziperman 2000).

We emphasize that, in the case of the uncoupled model, the system closely follows the thermal forcing up to the end of its increase (see Fig. 7 in Part I). This change in the behavior of the system is essentially due to the presence in the uncoupled model of a much stronger temperature-restoring coefficient, which summarizes the effects of all of the various surface heat fluxes.

In order to explore the processes responsible for the internal feedbacks characterizing the stability properties of the system around the initial state, we analyze in detail the diagnostics of the subcritical case, which is more instructive because the negative feedbacks eventually prevail on the external destabilizing perturbations.

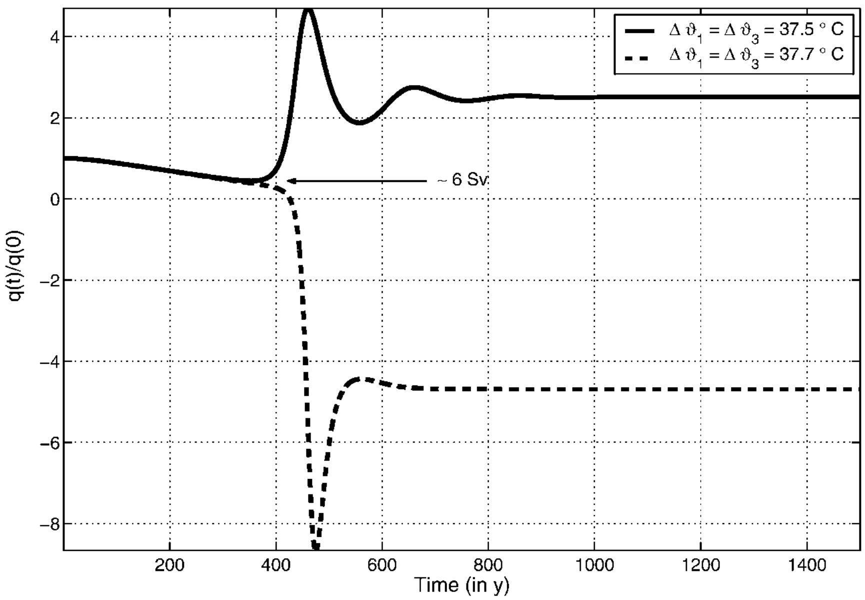

FIG. 1. Evolution of the THC strength under a super- and subcritical radiative flux forcing. In both cases $\Delta\vartheta_2 = 1.5\Delta\vartheta_1 = 1.5\Delta\vartheta_3$.

In Fig. 2 we analyze the effect of the radiative forcing on the meridional temperature gradient in parallel to the evolution of the THC. Figure 2b shows that both the Northern and Southern Hemisphere temperature gradients, after an initial slight increase due to the first-order effect related to the larger direct radiative forcing realized in the tropical regions, present a somewhat unexpected large decrease in response to the climate forcing, with the value of $T_2 - T_1$ reaching at equilibrium a value of ~10°C. The enhanced warming realized

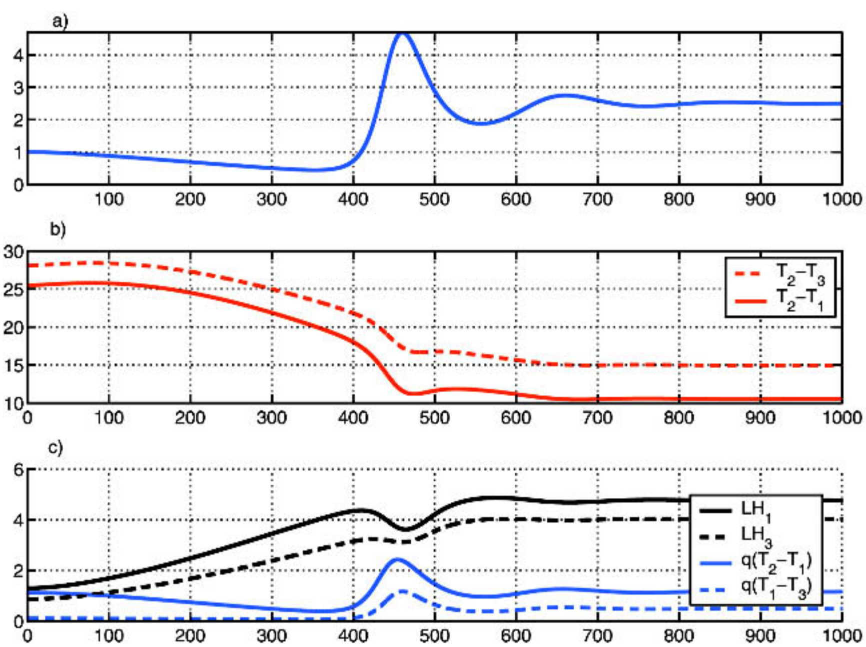

FIG. 2. (a) Evolution of the THC strength in units of $q(0)$ for the subcritical case presented in Fig. 1. (b) Evolution of the meridional temperature gradient $T_2 - T_1$ and $T_2 - T_3$ in units of °C. (c) Evolution of the latent heat fluxes and the oceanic advection thermal forcings $q(T_2 - T_1)$ (into box 1) and $q(T_1 - T_3)$ (into box 3) in units of $|\alpha q(0)\ [T_2(0) - 2T_1(0) + T_3(0)]|$. Abscissas show elapsed time expressed in y.

at the high latitudes qualitatively resembles the results of most GCMs in the simulation of global warming (Cubasch et al. 2001). Nevertheless, we need to point out that a major reason for high-latitude warming in these models is the albedo feedback, which we do not have. Moreover, in GCMs' transient experiments, they generally have strong high-latitude warming in the NH, but not in the SH, because of the stronger heat uptake in the SH ocean, which is not represented in our model (Cubasch et al. 2001). Figure 2c shows that in both hemispheres, the change of latent heat flux is the process providing the single most relevant contribution to this effect. While the forcing is dominant, the increase in the latent heat flux counteracts the decrease of the oceanic thermal advection caused by the decrease of the THC strength. At the final equilibrium, in both hemispheres the latent heat fluxes are much larger than the corresponding oceanic transports, while at the initial equilibrium the two processes are of comparable size. In the time frame where we observe the largest increase in the THC strength ($t \approx 400$ yr), we have a very fast decrease of the meridional temperature gradient caused by the great enhancement of the oceanic thermal advection. The sudden flattening of the meridional temperature structure causes a dip in the value of the latent heat fluxes.

Since, when a northern sinking equilibrium is realized, the value of the THC is monotonically increasing with the value of the freshwater flux in the upwelling box (Rahmstorf 1996; Scott et al. 1999; Part I), the increase of the value of F_3 (which is linear with LH_3) explains why we obtain at equilibrium a larger value for the THC strength. We wish to point out that the presence of an increase of the THC on a longer term agrees with some of the more complex models' results (Wang et al. 1999a; Wiebe and Weaver 1999).

In Fig. 3 we analyze the oceanic and atmospheric buoyancy forcings to the THC in parallel to the evolution of the THC. Such forcings are obtained as the difference between the buoyancy forcings in box 1 and 3. In Fig. 3b we present the evolution of the oceanic advective contributions to the tendency of the THC strength in units of the absolute value of the equilibrium thermal advective contribution. We observe that when the THC is forced to decrease, both the haline and thermal contributions counteract such change. While the negative advective thermal feedback is common in box models, the salinity advective feedback is usually positive. The peculiar behavior we observe for the haline part is because during the radiative forcing the moisture balance of the equatorial region is so negative that the northward salinity advection does not decrease. When the negative feedbacks of the system overcome the external forcing and cause a sudden large increase in the THC strength, the haline feedback and the thermal feedback get their usual signs, as can be observed in the spikelike structures for $t \approx 400$ yr. Moreover, after the end of the forcing, the thermal and haline contributions tend to be out of phase. In Fig. 3c we present the evolution of the atmospheric contributions to the forcings to the THC strength in the same units as in Fig. 3b. It is clear that the latent heat fluxes

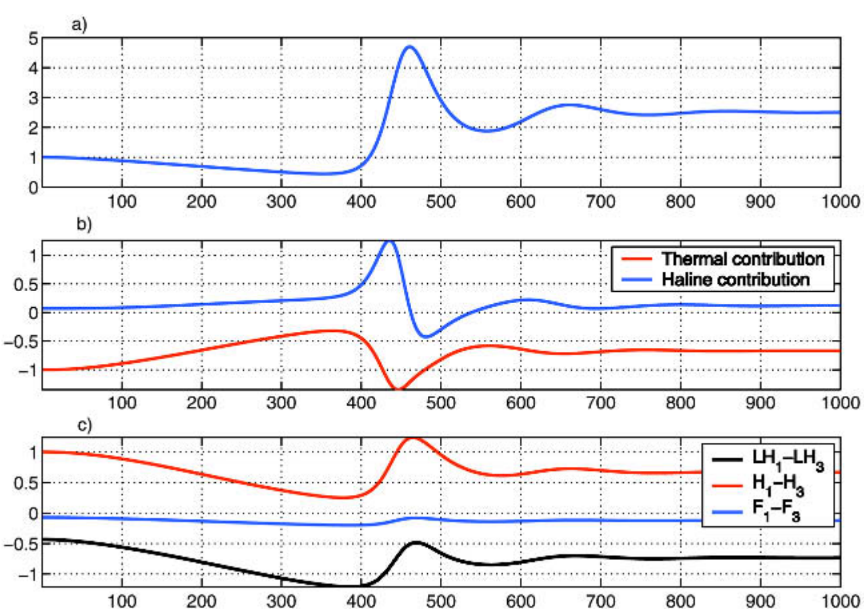

FIG. 3. (a) Evolution of the THC strength in units of $q(0)$ for the subcritical case presented in Fig. 1. (b) Evolution of the thermal and haline oceanic advection forcings to the strength of the THC in units of $|\alpha q(0)\ [T_2(0) - 2T_1(0) + T_3(0)]|$. (c) Evolution of the most relevant atmospheric contributions to the forcing to the strength of the THC [same units as in (b)]. Abscissas show elapsed time expressed in y.

play the most relevant role in the destabilization of the system for $t \leq 400$ yr since they contribute to all of the variations from the initial value of the atmospheric thermal forcing. The contribution related to the change of the freshwater fluxes is smaller than that provided by the changes of the latent heat fluxes by a factor of ~ 6 [see Eq. (7)]. Therefore, the largest contribution to the weakening of the THC is thermic and driven by the latent heat.

Two studies with coupled atmosphere–ocean GCMs have obtained contrasting results about the relative relevance of heat versus freshwater as destabilizing mechanisms; in Mikolajewicz and Voss (2000) it is shown that the heat flux change is the most important destabilizing mechanism and is dominated by the latent heat flux, just as in our model. By contrast, Dixon et al. (1999) conclude that changes in the moisture flux were the main destabilizing agent. However, calculations of the contributions of changes in both the heat and moisture fluxes to the change in the density flux in the model used in Dixon et al. (1999) show that it is in fact dominated by the heat flux changes (Huang et al. 2003). Indeed, this is true of all the seven coupled GCMs analyzed in the context of the Climate Intercomparison Project (CMIP) as shown by Huang et al. (2003). This apparent contradiction may be explained by the results presented in Kamenkovich et al. (2003), where it is found that, even though the decrease in the thermohaline circulation may be initiated by changes in the moisture flux, changes in the heat flux induced by atmospheric feedbacks nevertheless contribute strongly to the decrease.

We find that, in agreement with other studies considering either relatively sophisticated (Stocker and Schmittner 1997; Schmittner and Stocker 1999) or extremely simplified models (Tziperman and Gildor 2002), the evolution of latent heat and freshwater meridional fluxes under global warming scenarios depends on two competing effects: the reduced efficiency in the atmospheric transport due to decreased meridional temperature gradient and increased capability of the atmosphere to retain moisture for higher average temperatures. The second effect is thought to dominate for larger climate changes, as shown by paleoclimatic data and simulations of the last few hundred thousand years (Charles et al. 1994; Manabe and Stouffer 1994; Krinner and Genthon 1998; Kitoh et al. 2001), and it is reasonable to expect that similarly the Clausius–Clapeyron effect dominates the changes in the latent heat and freshwater fluxes in global warming scenarios (Tziperman and Gildor 2002).

We briefly analyze how sensitive the dynamics of the system is with respect to the choice of the eddy transport power law by imposing the same forcing as the subcritical case previously analyzed to the system obtained when the value of n is set to 3 and 5. In Fig. 4a we show the realized evolution of q in the three cases of $n = 1, 3, 5$. We see that the extent of the decline of the THC strength and its length are negatively correlated with the value of the exponent for the eddy transport power law. We have previously analyzed the prominent role of the changes in the latent heat fluxes in determining the dynamics of the system in the $n = 1$ case, so we limit the comparison to this process.

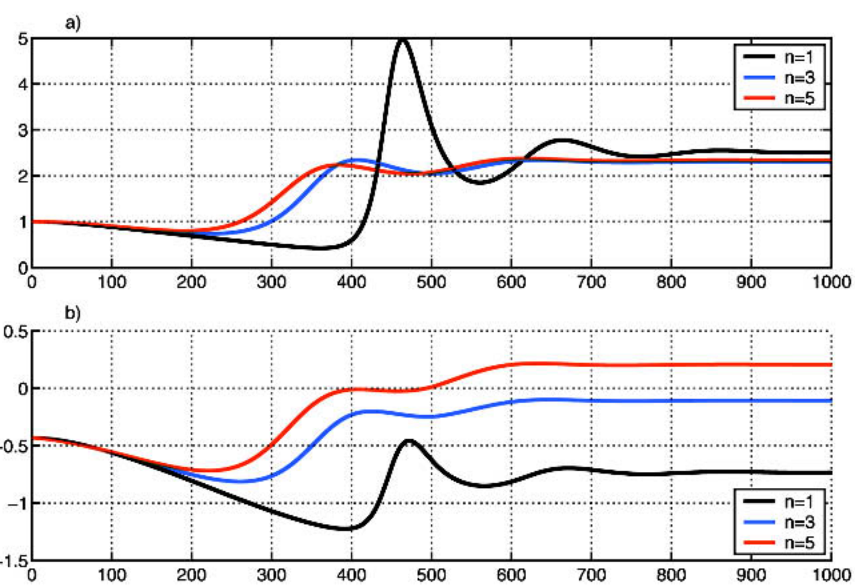

FIG. 4. (a) Evolution of the THC strength in units of $q(0)$ for various values of n. Radiative forcing is the same as the subcritical case presented in Fig. 1. (b) Evolution of the latent heat flux contributions $LH_1 - LH_3$ to the forcing for various values of n in units of $|\alpha q(0)\ [T_2(0) - 2T_1(0) + T_3(0)]|$. Abscissas show elapsed time expressed in y.

We wish to emphasize that in the $n = 3$ and $n = 5$ cases we observe the decrease in the meridional temperature gradient in spite of a stronger tropical forcing (not shown). Analogously, we note that in all cases analyzed the final value of the THC is larger than the initial value, and that this positively correlates with the positive difference between the final and initial values of LH_2 (not shown). Therefore, such features seem robust.

In Fig. 4b we show that the behavior of the latent heat contribution changes with n coherently with the previously presented picture. We show that the timing of the *minima* of the latent heat contributions coincide in all cases with the timing of the THC-strength minima, and the same applies for the *maxima*. Moreover, comparing the three cases, the minima of the latent heat contributions are larger in absolute value when the corresponding minima of the THC strength are deeper. Our results suggest that a more temperature-gradient-sensitive atmospheric transport is more effective in limiting the destabilizing processes, in which the changes in the latent heat fluxes are the most prominent.

5. Hysteresis

Choosing quasi-static perturbations to the radiative temperatures can lead to the reversal of the THC only if we select $n = 1$, the main reason being that if the atmospheric transport is more sensitive to the meridional temperature gradient, the system is always able to counteract a slowly increasing destabilizing radiative forcing. This qualitative difference between the behavior of the various versions of the model is independent of the selected value of the parameter TRRF. With the choice of $n = 1$ and TRRF = 1.5, quasi-static perturbations are destabilizing only if $\Delta\vartheta_3/\Delta\vartheta_1 \leq 0.5$. We present in Fig. 5 the hysteresis graphs relative to $\Delta\vartheta_3/\Delta\vartheta_1 = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$. In Fig. 5a we have $\Delta\vartheta_1$ as abscissa and q in units of the initial equilibrium value as ordinate, so that the initial state is the point (0, 1). The abscissas of the bifurcation points on the right-hand side of the graph increase with increasing value of $\Delta\vartheta_3/\Delta\vartheta_1$ from ~12° to ~25°C, while the ordinates are close to 1, thus implying that when the bifurcation occurs the value of the THC is only slightly different from the initial equilibrium value. The bistable region is remarkably large in all cases, the total extent increasing with increasing values of $\Delta\vartheta_3/\Delta\vartheta_1$, because the abscissas of the bifurcation points in the left-hand side of the graph circuits are in all cases below −30°C and so well within the unphysical region of parameter space. This means that once the circulation has reversed, the newly established pattern is extremely stable and can hardly be changed again. In Fig. 5b we present the corresponding hysteresis graphs where the abscissa, in this case, is the value of the realized freshwater flux into box 1 when the radiatively forced system has reached a newly established equilibrium. The initial equilibrium is the point (1, 1); we can see that there is a monotonic 1 to 1 mapping between Figs. 5a and 5b, apart from a very limited region around the bifurcation point in the right-hand side for low values of $\Delta\vartheta_3/\Delta\vartheta_1$. This implies that

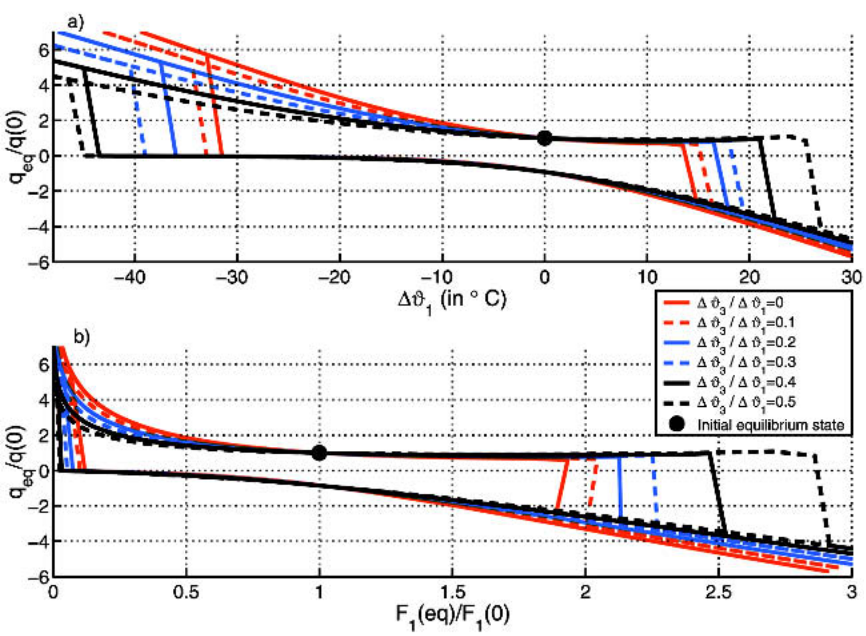

FIG. 5. Hysteresis graphs of the system with $n = 1$ for radiative temperature perturbations. (a) Radiative temperature change $\Delta\vartheta_1$ is considered as the relevant parameter. (b) Freshwater flux F_1 is considered as the relevant parameter.

in our model, at equilibrium, the changes of the average surface temperature and the changes in the hydrological cycle are positively correlated, as one would expect. We observe that in terms of freshwater flux, the bistable region is somewhat smaller than the uncoupled case (see Fig. 5 in Part I), thus suggesting a caveat in the interpretation of uncoupled models' results.

6. Critical perturbations

In this section we present the set of critical forcings that divide the forcings that disrupt the present pattern of the THC from those that drive the system to a northern sinking state qualitatively similar to the initial unperturbed state. We extend the analysis of the hysteresis of the system by considering how the temporal pattern of the perturbation influences the ability of the system to counteract destabilizing forcings. In Fig. 6 we consider the model plus forcing case characterized by $n = 1$ and TRRF = 1.5. In Fig. 6a we present the manifold of those critical forcings using the coordinate system $[t_0, \Delta\vartheta_3/\Delta\vartheta_1, \Delta\vartheta_1]$, while in Fig. 6b we adopt the coordinate system $[t_0, \Delta\vartheta_3/\Delta\vartheta_1, \vartheta_1']$. We present the corresponding results for $n = 3$ and $n = 5$ in Figs. 7 and 8, respectively. We adopt in all cases a logarithmic scale since, acknowledging the limitations of our model, we are mainly interested in capturing the qualitative properties of the response of the system. There is a general agreement between the response of the coupled and of the uncoupled model (Part I) to destabilizing perturbations. Figures 6–8 all suggest that the more symmetric and slower the forcing, the less likely is the destabilization of the THC:

- for a given t_0, the lower the value of the ratio $\Delta\vartheta_3/\Delta\vartheta_1$, the lower the total change $\Delta\vartheta_1$ needed to obtain the reversal of the THC;
- for a given value of the ratio $\Delta\vartheta_3/\Delta\vartheta_1$, more rapidly increasing perturbations (larger ϑ_1') are more effective in disrupting the circulation.

We need to point out a very relevant feature that shows how critical the choice of the parameter n is. In the $n = 1$ case, consistent with the study of the hysteresis runs presented in Fig. 5, very slow perturbations can cause the collapse of the THC for low values of $\Delta\vartheta_3/\Delta\vartheta_1$. On the contrary, in the $n = 3$ and $n = 5$ cases there is, even for very low values of $\Delta\vartheta_3/\Delta\vartheta_1$, a threshold in the rate of increase of the forcing below which the reversed THC does not occur, independent of the total radiative forcing realized. This can be more clearly understood by observing that in Figs. 7b and 8b, for each value of $\Delta\vartheta_3/\Delta\vartheta_1$, the critical value of the rate of increase of the forcing ϑ_1' is independent of the temporal extension of the forcing t_0 for large values of $t_0 \geq t_s$. In the $n = 3$ and $n = 5$ cases the system cannot make transitions to a southern sinking equilibrium for quasi-static perturbations, since they would require indefinitely large perturbations.

Previous studies focusing on more complex models obtain a similar dependence of thresholds on the rate of increase of the forcings (Stocker and Schmittner 1997; Schmittner and Stocker 1999), while in other studies

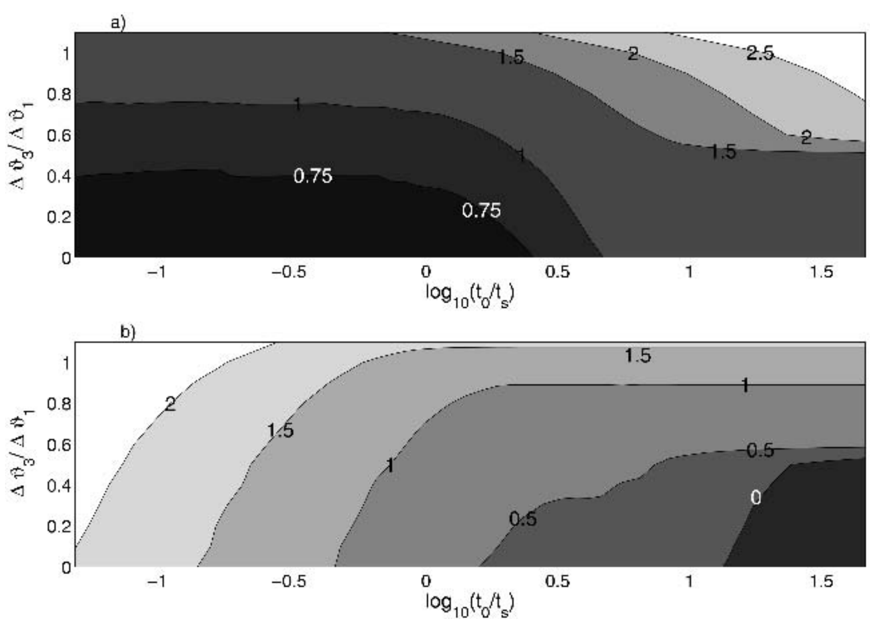

FIG. 6. Stability of the THC against perturbations to the radiative temperature with $n = 1$. (a) Critical values of the total increase of the freshwater flux; contours of $\log_{10}[\Delta\vartheta_1]$, where $\Delta\vartheta_1$ is in °C. (b) Critical values of the rate of increase of the freshwater flux. Contours are of $\log_{10}[\vartheta_1']$, where ϑ_1' is in °C t_s^{-1} and t_s is the advective time.

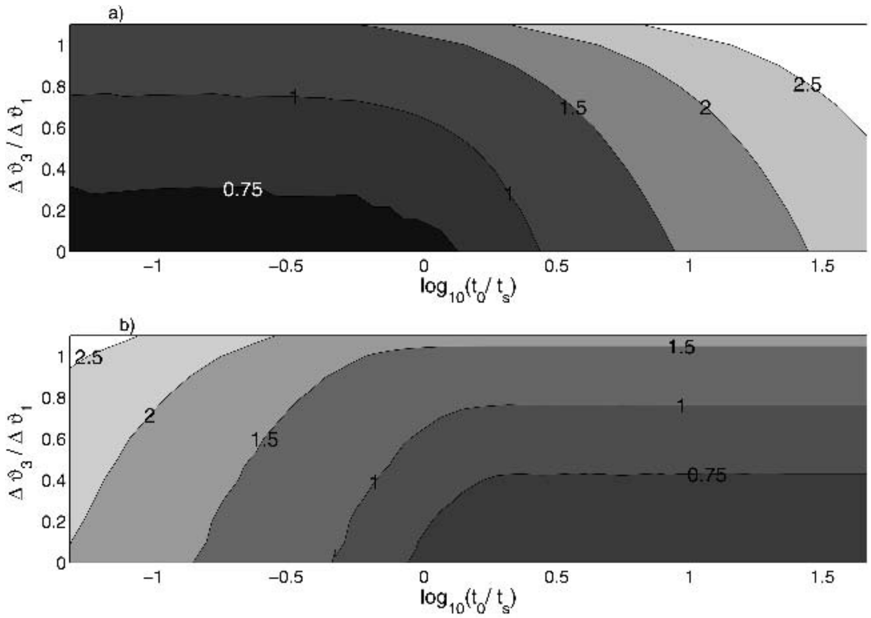

FIG. 7. Stability of the THC against perturbations to the radiative temperature with $n = 3$. (a) Critical values of the total increase of the freshwater flux; contours are of $\log_{10}[\Delta\vartheta_1]$, where $\Delta\vartheta_1$ is in °C. (b) Critical values of the rate of increase of the freshwater flux; contours are of $\log_{10}[\vartheta_1']$, where ϑ_1' is in °C t_s^{-1} and t_s is the advective time.

where the full collapse of the THC is not obtained, it is nevertheless observed that the higher the rate of increase of the forcing, the larger the decrease of the THC realized (Stouffer and Manabe 1999). Other studies on coupled models also show how the spatial pattern of freshwater forcing due to global warming is extremely relevant especially in the short time scales; only forcings occurring mainly in the northern Atlantic are

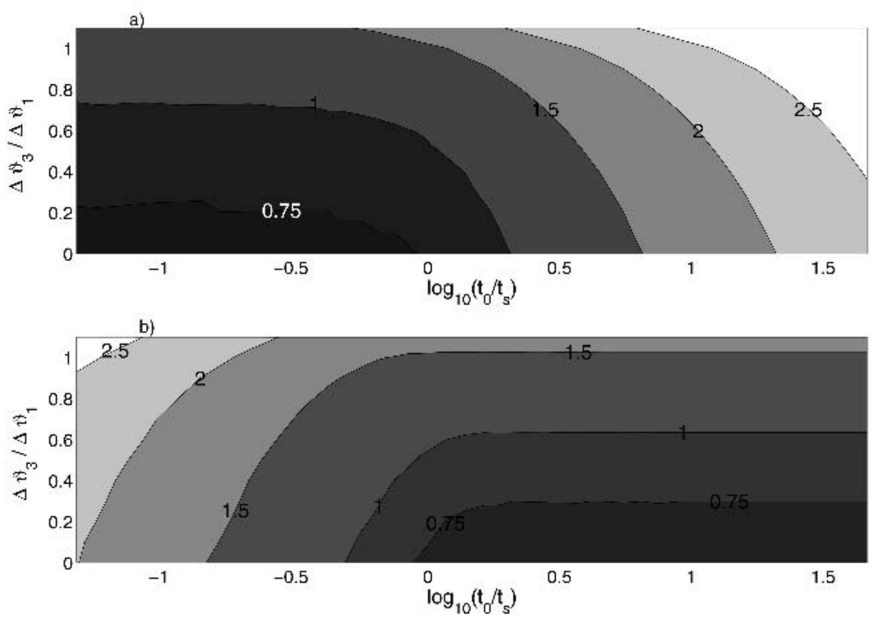

FIG. 8. Stability of the THC against perturbations to the radiative temperature with $n = 5$. (a) Critical values of the total increase of the freshwater flux; contours are of $\log_{10}[\Delta\vartheta_1]$, where $\Delta\vartheta_1$ is in °C. (b) Critical values of the rate of increase of the freshwater flux; contours are of $\log_{10}[\vartheta_1']$, where ϑ_1' is in °C t_s^{-1} and t_s is the advective time.

efficient in destabilizing the THC (Rahmstorf 1996; Rahmstorf and Ganopolski 1999; Manabe and Stouffer 2000; Ganopolski et al. 2001).

7. Sensitivity study

From the analysis of the Figs. 1–4 we have hinted that the difference between the latent heat fluxes into the two high-latitude boxes dominates the dynamics of the forced system and determines its stability. Given the properties and the functional form of LH_1 and LH_3 we expect that

1) the system is less stable against radiative forcings if the tropical-to-high-latitude radiation forcing ratio is larger, and
2) in the case when the radiative forcing is larger in the northern high-latitude box, the system is more stable if the atmospheric transport feedback is stronger; this effect is likely to be notable *only if* the perturbations have time scales larger than the flushing time of the oceanic boxes, as confirmed in the extreme case of quasi-static perturbations.

To obtain a more quantitative measure of which processes are important, we analyze the sensitivity of the vertical coordinate of the the manifold of the critical perturbations shown in Fig. 6 to changes in the two key parameters, TRRF and n. We define Z_C (t_0, $\Delta\vartheta_3$, *TRRF*, n) as the function giving the critical value of the change in the radiative temperature of box 1, which was thoroughly described in the previous section.

Figure 9 shows the value of the two-dimensional field of the finite differences $\log_{10}$ [Z_C (t_0, $\Delta\vartheta_3$, TRRF = 1.0, n = 3)] − $\log_{10}$ [Z_C (t_0, $\Delta\vartheta_3$, TRRF = 2.0, n = 3)]. We observe that the field does not remarkably depend on the perturbation length t_0; this suggests that processes controlling the dynamics of the systems acting on very different time scales are similarly affected by changes in TRRF. This means that the increase in the ratio between the tropical and the northern high-latitude radiative forcing changes the response of the system and favors the collapse of the THC evenly and independently of the temporal scale of the forcing itself, and is particularly effective if we use quasi-symmetric forcings. Therefore the change of the parameter TRRF = $\Delta\vartheta_2/\Delta\vartheta_1$ does not have preferential effect on any of the feedbacks.

Figure 10 presents the value of the two-dimensional field of the finite differences $\log_{10}$ [Z_C (t_0, $\Delta\vartheta_3$, TRRF = 1.5, n = 5)] − $\log_{10}$ [Z_C (t_0, $\Delta\vartheta_3$, TRRF = 1.5, n = 3)]. The most striking feature of this field is that there is a very strong gradient only along the t_0 direction and for $t_0 \approx t_s$. The field is small and positive for forcings having a temporal scale shorter than the characteristic oceanic time scale of the system, while it becomes very large and positive for forcings having long temporal scales. The positive value means that a more sensitive atmospheric transport (higher values of n) tends to stabilize the system if $\Delta\vartheta_3/\Delta\vartheta_1 \leq 1$. With a more temperature-gradient-sensitive atmospheric transport, smaller changes in temperature gradients between the boxes are needed (see Fig. 2) to change all the atmospheric fluxes; therefore the system is able to dynamically arrange the atmospheric fluxes very effectively, so that

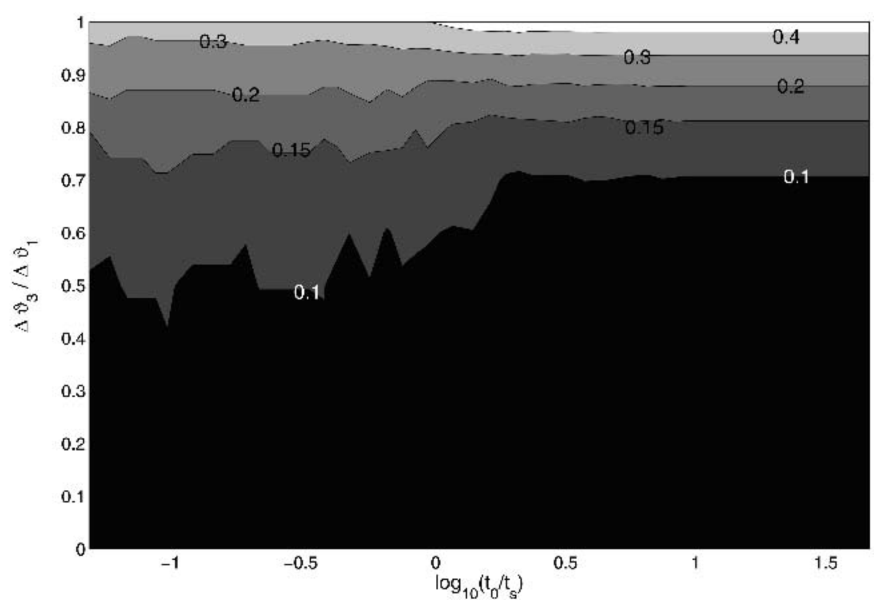

FIG. 9. Sensitivity to the low-to-high-latitude radiative forcing ratio TRRF of the critical values of the total increase of the radiative temperatures; contours are of $\log_{10}$[Z_C (TRRF = 1.0) / Z_C (TRRF = 2.0)].

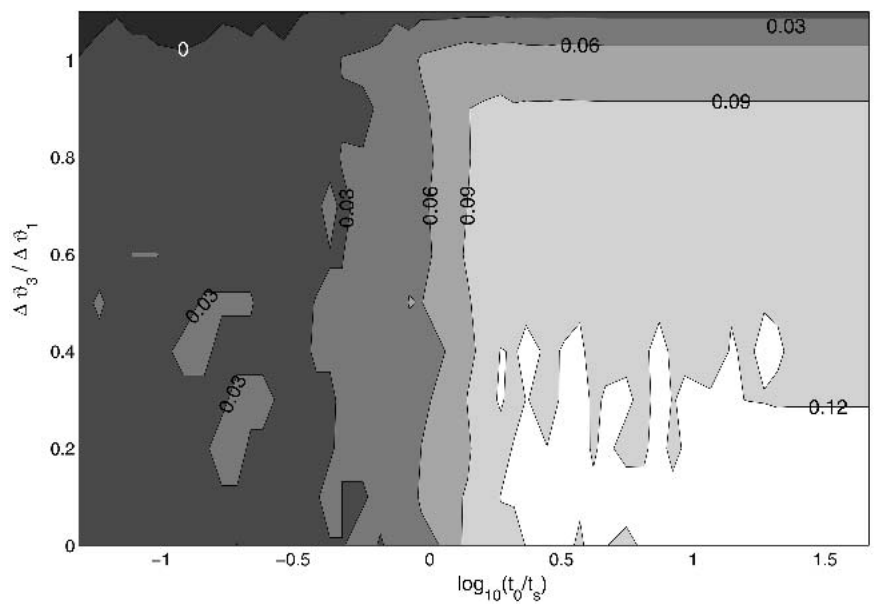

FIG. 10. Sensitivity to the atmospheric transport parameterization of the critical values of the total increase of the radiative temperatures; contours are of $\log_{10}[Z_C\ (n = 5) / Z_C\ (n = 3)]$.

they can more efficiently counteract the external forcings and keep the system as close as possible to the initial state. If the forcings are very fast, the enhancement of the atmospheric stabilizing mechanism is not very effective. On the contrary, for slow forcings involving time scales comparable to or larger than those of the system, the enhanced strength of the negative feedback obtained with the increasing efficiency of the atmospheric transport can play a very significant stabilizing role in the dynamics of the system. We have chosen the $n = 5$ and $n = 3$ difference field, because it gives a clearer picture of the phenomenon we wish to emphasize. The previous observations provide an explanation of why the case $n = 1$ is more unstable especially for slow forcings.

These conclusions seem to be in contrast to the result that in several uncoupled models, shorter relaxation times for box temperatures—which, as shown by Marotzke (1996), correspond to more sensitive atmospheric heat transports—imply less stability for the system (Tziperman et al. 1994; Nakamura et al. 1994; Scott et al. 1999; Rahmstorf 2000). Actually, the contrast is only apparent because in an uncoupled model there is no adjustment of moisture fluxes due to temperature changes; therefore decreasing the relaxation time makes the system less flexible and adaptable to forcings. Moreover, our results seem to disagree with the conclusions drawn in the coupled model presented by Scott et al. (1999). We think this disagreement is because in our study we are dealing with a different kind of perturbation, descriptive of global warming, and that these perturbations have a direct and strong influence on the most relevant atmospheric fluxes because of their highly nonlinear dependence on average temperature changes. This property was not present in the parameterizations used by Scott et al. (1999).

By contrast with the above-described result, we see in Fig. 10 that for $\Delta\vartheta_3/\Delta\vartheta_1 > 1$ and $t_0 \leq 80$ yr, the difference field changes sign. This implies that for extremely fast radiative forcings that are larger in box 3, a system with larger n is more easily destabilized. This occurs because for such forcings the northern meridional temperature gradient is initially forced to increase more than the southern, so that the northward atmospheric fluxes are greatly enhanced. Such a process tends to destabilize the circulation and is stronger for larger values of n. As observed in Part I, since we have considered the approximation of well-mixed boxes and excluded from the model the equatorial deep box used by other authors (Rahmstorf 1996), our model should not be considered very reliable for phenomena taking place in time scales $\ll t_s$.

8. Conclusions

In this paper we have analyzed the stability of the THC as described by a set of coupled models differing in the ratio between the radiative forcing realized at the Tropics and high latitudes and in the parameterization of the atmospheric transports.

In a coupled model a natural representation of the radiative forcing is possible, since the main atmospheric physical processes responsible for freshwater and heat fluxes are formulated separately. Although only weakly asymmetric or symmetric radiative forcings are representative of physically reasonable conditions, we have considered general asymmetric forcings in order to get a more complete picture of the mathematical properties of the system.

We have analyzed five combinations of the model plus radiative forcing, considering different combinations of the atmospheric transport parameterization and of the ratio between the high-to-low-latitude radiative forcing.

When the system is radiatively forced, initially the latent heat fluxes and the freshwater fluxes are strongly enhanced thanks to the increase in the saturation pressure of the water vapor due to the warming, and the fluxes into the northern high-latitude box increase. These are the main causes for the initial reduction of the THC strength. The strong increase of heat flux into box 1 eventually reduces the efficacy of the atmospheric transport and so causes a great reduction of the freshwater flux and the latent heat (and sensible heat) flux into box 1, which induces an increase in the THC. When the radiative forcings do not drive the system to a collapse, the enhancement of meridional heat fluxes causes large reductions in the meridional temperature gradients, even when larger radiative forcings are prescribed at the Tropics. We note that such a behavior has been obtained even though our model does not include the ice–albedo feedback. This result seems relevant in the interpretation of warm paleoclimates, such as during the Eocene.

The variations of latent heat and freshwater fluxes into the two high-latitude boxes dominate the dynamics of the forced coupled model. This is because of the very strong nonlinear dependence of water vapor saturation pressure on temperature. Therefore the inclusion of the Clausius–Clapeyron relation in the fluxes' parameterization seems to play a key role in all the results obtained in our study. In our system the major role is played by the latent heat because in density terms it is stronger than the freshwater flux by a factor of $\sim$6. A qualitatively similar dominance is found in the CMIP models (Huang et al. 2003; Mikolajewicz and Voss 2000), while other studies give the opposite result (Stocker and Schmittner 1997; Rahmstorf and Ganopolski 1999; Schmittner and Stocker 1999; Manabe and Stouffer 1999b; Ganopolski et al. 2001).

The hysteresis graphs of the system show formally that that quasi-static perturbations cannot disrupt the northern sinking pattern of the circulation unless we are considering a relatively inefficient atmospheric transport and radiative forcing with large, unrealistic north–south asymmetry. Moreover, in this case the initial equilibrium point is in the bistable region.

We obtain, with a parametric study involving the total forcing realized, its rate of increase, and its north–south asymmetry, the manifold of the critical forcings dividing the forcings driving the system to a southern sinking equilibrium from those that do not qualitatively change the pattern of the THC. We generally find that fast forcings are more effective than slow forcings in disrupting the present THC patterns, forcings that are stronger on the northern box are also more effective in destabilizing the system, and very slow forcings do not destabilize the system whatever their asymmetry, unless the atmospheric transport is only weakly dependent on the meridional temperature gradient.

We also compute the sensitivity of the results obtained with respect to the tropical-to-high-latitude radiative forcing and the efficiency of the atmospheric transport. A higher forcing in the Tropics destabilizes the system evenly at every time scale and greatly favors destabilization for quasi-symmetric forcings. Increasing the efficiency of the atmospheric transports makes the system in general more stable against destabilizing radiative forcings because it allows a very effective control of all the buoyancy fluxes and provides an enhancement of the atmospheric transport's negative feedback. This effect is especially relevant for forcings' time scales comparable or larger than the system's characteristic time scale, while it is not notable for fast forcings that bypass all the feedbacks of the system.

The adoption of a linear diffusive representation of the atmospheric transport greatly affects the main qualitative aspects of the stability properties of the system. Therefore, one should take great care in adopting such a questionable (Stone and Miller 1980; Stone and Yao 1990) parameterization in studies where large climatic shift are investigated.

Comparing the results obtained in this study with those presented in Part I, we conclude that the introduction of an explicit atmosphere–ocean coupling increases the stability of the system so that obtaining the collapse of the THC requires very extreme forcings. We conclude that the coupling introduces a negative net feedback, which is stronger with more efficient atmospheric transports. We can guess that similar effects can be observed with more complex models, thus introducing a caveat in the interpretation of data coming from uncoupled oceanic models.

The relevance of the temporal scale of the forcing in determining the response of our system to perturbations affecting the stability of the THC confirms the findings of Tziperman and Gildor (2002) for a hemispheric coupled box model, of Stocker and Schmittner (1997) and Schmittner and Stocker (1999) in the context of Earth system models of intermediate complexity (EMICs), and of Manabe and Stouffer (1999a,b, 2000), and Stouffer and Manabe (1999) in the context of GCMs. The coherence within the whole hierarchical ladder of models gives robustness to this result.

We conclude that when analyzing the behavior of the THC in global warming scenarios with more complex coupled models, the spatial pattern of the radiative forcing should be carefully taken into account because the THC is a highly nonlinear, nonsymmetric system, and the effect of changing the rate of increase of the radiative forcing should be explored in great detail.

This would provide a bridge between the instantaneous and quasi-static changes. The exploration of the THC dynamics in global warming scenarios requires such an approach since the system encompasses very different time scales that can be explored only if the timing of the forcing is varied. Moreover, the sensitivity of our system's response to the parameter n suggests that, in order to capture the behavior of the THC in the context of relevant changes of the climate, it is critical for coupled models to capture precisely the processes responsible for fueling the atmospheric transport.

This work allows many improvements, some of which are similar to those proposed in Part I. In particular, asymmetries in the oceanic fractional areas would induce asymmetries in the values of B_i, thus causing the presence of different restoring times for the various boxes, while asymmetries in the freshwater catchment areas ($\gamma_1 \neq \gamma_3$) would make the relative importance of the latent heat and freshwater fluxes (when expressed in common density units) different in the two boxes $i = 1, 3$.

The presence of the albedo feedback would also enhance any asymmetry between the two high-latitude boxes; it could be included in the model by parameterizing the radiative terms A_i as increasing functions of the temperatures T_i, along the lines of Stocker and Schmittner (1997), Schmittner and Stocker (1999), and Tziperman and Gildor (2002), considering the temperatures as proxies for the fraction of the surface covered by ice. We expect that the inclusion of an ice–albedo feedback would decrease the stability of the THC.

The model's reliability would also benefit from the inclusion of a more appropriate nonlinear equation of state for the seawater. A likely effect would be increasing the relative relevance of the freshwater flux changes as a driving mechanism of the weakening of the THC.

Acknowledgments. The authors are grateful to J. Scott for interesting discussions and useful suggestions. One author (V.L.) wishes to thank R. Stouffer, for having proposed improvements to an earlier version of the manuscript, and T. Stocker and E. Tziperman for having suggested a number of relevant references. This research was funded in part by the U.S. Department of Energy's (DOE's) Climate Change Prediction Program and in part by the MIT Joint Program on the Science and Policy of Global Change (JPSPGC). Financial support does not constitute an endorsement by DOE or JPSPGC of the views expressed in this article.

REFERENCES

Charles, C. D., D. Rind, J. Jouzel, R. D. Koster, and R. G. Fairbanks, 1994: Glacial/interglacial changes in moisture sources for Greenland: Influences on the ice core record of climate. *Science,* **263,** 508–518.

Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton, Eds., Cambridge University Press, 526–582.

Dixon, K. W., T. L. Delworth, M. J. Spelman, and R. J. Stouffer, 1999: The influence of transient surface fluxes on North Atlantic overturning in a coupled GCM climate change experiment. *Geophys. Res. Lett.,* **26,** 2749–2752.

Ganopolski, A., V. Petoukhov, S. Rahmstorf, V. Brovkin, M. Claussen, and C. Kubatzki, 2001: CLIMBER-2: A climate system model of intermediate complexity. Part II: Model sensitivity. *Climate Dyn.,* **17,** 735–751.

Held, I. M., 1978: The vertical scale of an unstable baroclinic wave and its importance for eddy heat flux parameterizations. *J. Atmos. Sci.,* **35,** 572–576.

Huang, B., P. H. Stone, A. P. Sokolov, and I. V. Kamenkovich, 2003: The deep-ocean heat uptake in transient climate change. *J. Climate,* **16,** 1352–1363.

Kamenkovich, I. V., A. P. Sokolov, and P. H. Stone, 2003: Feedbacks affecting the response of the thermohaline circulation to increasing CO_2: A study with a model of intermediate complexity. *Climate Dyn.,* **21,** 119–130.

Kitoh, A., S. Murakami, and H. Koide, 2001: A simulation of the Last Glacial Maximum with a coupled atmosphere–ocean GCM. *Geophys. Res. Lett.,* **28,** 2221–2224.

Krinner, G., and C. Genthon, 1998: GCM simulations of the Last Glacial Maximum surface climate of Greenland and Antarctica. *Climate Dyn.,* **14,** 741–758.

Lucarini, V., and P. H. Stone, 2005: Thermohaline circulation stability: A box model study. Part I: Uncoupled model. *J. Climate,* **18,** 501–513.

Manabe, S., and R. J. Stouffer, 1994: Multiple-century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Climate,* **7,** 5–23.

——, and ——, 1999a: Are two modes of thermohaline circulation stable? *Tellus,* **51A,** 400–410.

——, and ——, 1999b: The role of thermohaline circulation in climate. *Tellus,* **51A,** 91–109.

——, and ——, 2000: Study of abrupt climate change by a coupled ocean–atmosphere model. *Quat. Sci. Rev.,* **19,** 285–299.

Marotzke, J., 1996: Analysis of thermohaline feedbacks. *Decadal Climate Variability: Dynamics and Predicatibility,* D. L. T. Anderson and J. Willebrand, Eds., NATO ASI Series, Vol. 44, Springer, 333–378.

——, and P. H. Stone, 1995: Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *J. Phys. Oceanogr.,* **25,** 1350–1360.

Mikolajewicz, U., and R. Voss, 2000: The role of the individual air–sea flux components in CO_2-induced changes of the ocean's circulation and climate. *Climate Dyn.,* **16,** 627–642.

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

Peixoto, A., and B. Oort, 1992: *Physics of Climate.* American Institute of Physics, 520 pp.

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

——, 2000: The thermohaline ocean circulation—A system with dangerous thresholds? *Climate Change,* **46,** 247–256.

——, and A. Ganopolski, 1999: Long-term global warming scenarios computed with an efficient coupled climate model. *Climate Change,* **43,** 353–367.

Ramanathan, V., M. S. Lian, and R. D. Cess, 1979: Increased atmospheric CO_2: Zonal and seasonal estimates of the effect on the radiation energy balance and surface temperature. *J. Geophys. Res.,* **84,** 4949–4958.

Schmittner, A., and T. F. Stocker, 1999: The stability of the thermohaline circulation in global warming experiments. *J. Climate,* **12,** 1117–1127.

Scott, J. R., J. Marotzke, and P. H. Stone, 1999: Interhemispheric thermohaline circulation in a coupled box model. *J. Phys. Oceanogr.,* **29,** 351–365.

Shine, K. P., and Coauthors, 1995: Radiative forcing due to changes in ozone: A comparison of different codes. *Atmospheric Ozone as a Climate Gas,* W.-C. Wang and I. S. A. Isaksen, NATO ASI Series, Vol. 32, Springer, 373–396.

Stocker, T. F., and A. Schmittner, 1997: Influence of CO_2 emission rates on the stability of the thermohaline circulation. *Nature,* **388,** 862–864.

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

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

Stouffer, R. J., and S. Manabe, 1999: Response of a coupled ocean–atmosphere model to increasing atmospheric carbon dioxide: Sensitivity to the rate of increase. *J. Climate,* **12,** 2224–2237.

Tziperman, E., 2000: Uncertainties in thermohaline circulation response to green-house warming. *Geophys. Res. Lett.,* **27,** 3077–3080.

——, and H. Gildor, 2002: The stabilization of the thermohaline circulation by the temperature–precipitation feedback. *J. Phys. Oceanogr.,* **32,** 2707–2714.

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

Wang, W. C., and P. H. Stone, 1980: Effect of ice–albedo feedback on global sensitivity in a one-dimensional radiative–convective climate model. *J. Atmos. Sci.,* **37,** 545–552.

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

——, ——, and ——, 1999b: Thermohaline circulation. Part II: Sensitivity with interactive atmospheric transport. *J. Climate,* **12,** 83–92.

Wiebe, E. C., and A. J. Weaver, 1999: On the sensitivity of global warming experiments to the parametrisation of sub-grid scale ocean mixing. *Climate Dyn.,* **15,** 875–893.