Variability and Multiple Equilibria of the Thermohaline Circulation Associated with Deep-Water Formation

G. LENDERINK AND R. J. HAARSMA

KNMI, De Bilt, the Netherlands

(Manuscript received 15 April 1993, in final form 21 September 1993)

ABSTRACT

In this study the variability of the thermohaline circulation on decadal and centennial time scales that is related to the process of deep-water formation is investigated. This is done within the context of a simple geostrophic three-layer ocean model with a rectangular closed basin. When slowly varying the atmospheric forcing the model response shows sudden transitions, characterized by local changes in convective activity. Many different equilibria were found within the thermally driven regime (in Stommel's sense, i.e., downwelling occurs near the poles). Next, the deep-water formation process was analyzed with a one-dimensional box model. In this box model four different regimes can be identified: convective, nonconvective, periodic, and a regime where both convection and no convection are possible with the same mixed boundary conditions. These regimes were identified in the circulation of the ocean model. A region was traced where convection is possible according to the authors' analysis but not occurring. In this potentially convective region convection can be triggered easily. Similarly, there exists a region where the deep-water formation is easily suppressed. These sensitive areas generate multiple equilibria in the ocean model and contain a simple mechanism for variability on decadal and centennial time scales in the climate system.

1. Introduction

Recently, the process of deep-water formation has received much attention from climate modelers. The large-scale meridional ocean circulation and the process of deep-water formation are inextricably intertwined. At present the main source of deep water is North Atlantic Deep Water (NADW). Associated with NADW formation is the North Atlantic Current, to which northwestern Europe owes its relatively mild climate.

There has been much speculation about the existence of multiple equilibria of the (atmosphere–) ocean system. Stommel (1961) showed with a simple box model that the thermohaline circulation may have two significantly different modes. The thermally driven (normal) mode has deep-water formation near the pole; the salt driven (reversed) mode has deep-water formation near the equator. With respect to this bimodality much work on the thermohaline circulation has been undertaken. Broecker et al. (1985) suggested that the oceanic system exhibits two stable regimes of flow, respectively, characterized by the presence and absence of NADW formation. In their opinion the no-NADW circulation corresponds to the glacial mode. Duplessy et al. (1988) proposed that the deep-water formation in the North Atlantic during the last glacial maximum was much reduced; the main source of deep water being near Antarctica. At present, however, there is still little evidence that deep water has been formed near the equator during recent history.

Recent studies with ocean general circulation models (OGCMs) confirm the significance of deep-water formation for the thermohaline circulation. Deep-water formation, parameterized by the convective adjustment procedure, plays a crucial role in the simulation of the ocean circulation. Bryan (1986a,b) found that an interruption of the deep-water formation, either spontaneously or by a small salt anomaly, could lead to a temporal reversal of the circulation, which he called the polar halocline catastrophe. Due to a steady warming of the deep ocean during the reversed circulation, eventually a static instability occurs in subpolar regions. This leads to extremely strong deep-water formation, and a corresponding strong thermally driven circulation, called a flush. The occurrence of flushes and polar catastrophes has been confirmed by many other authors in different models (e.g., Marotzke 1990; Wright and Stocker 1991). Weaver and Sarachik (1991) found that the advection of positive salinity anomalies from the subpolar regions, where they are generated, to the polar regions, where they influence the deep-water formation, could lead to decadal oscillations. Multiple stable steady states were found by Marotzke and Willebrand (1991) for a highly idealized model of the World Ocean. These multiple equilibria were also found by Hughes and Weaver (1994) and Power and Kleeman (1994) for a

Corresponding author address: Geert Lenderink, Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, the Netherlands.

© 1994 American Meteorological Society

more realistic ocean basin. Two stable steady states in a coupled ocean–atmosphere model were found by Manabe and Stouffer (1988). The place and strength of deep-water formation of the equilibria, found in these studies, differ.

Until now, modeling studies were mostly focused on the dominant modes of the thermohaline circulation. The transitions between these modes are thought to be associated with large climatic changes, such as the transition between glacial and nonglacial episodes. Recently, investigators have tried to estimate from deep sea cores (e.g., $\delta^{13}C$ and $\delta^{18}O$ isotopes) the fractions of deep water originating from the various regions of deep-water formation during the last hundred thousand years. The results of these studies suggest that the deep ocean circulation exhibits many different states (Oppo and Fairbanks 1990) or even a continuum of circulation modes (Raymo et al. 1990) instead of a simple bimodality. Even more recently, these views have been supported by Lehman and Keigwin (1992) and Veum et al. (1992). Changes between these modes are rather abrupt. Differences between the states are characterized by the place, strength, and penetration depth of the deep-water formation. They are not as large as those involved with a simple bimodality but are nevertheless significant for variations of climate.

These studies give rise to a few questions. First, the possible equilibria of the thermohaline circulation must be established. Second and perhaps even more important, the stability of those equilibria and the transition process between those equilibria must be understood. Several studies tried to tackle the first question; the second question, however, remains largely unanswered. Stability of an equilibrium was always determined by large-scale perturbations. It was never attempted, according to our knowledge, to determine more precisely in which regions the thermohaline circulation is sensitive to perturbations. We tried to get more insight into those dynamics by considering a simple ocean model consisting of three homogeneous layers in the vertical. Assuming only latitude-dependent distributions for the atmospheric forcing, we show that many different equilibria exist within the thermally driven regime under identical mixed boundary conditions. These mixed boundary conditions are identical in a mathematical sense; physically they represent a very crude coupling to a constant atmosphere with infinite heat capacity. The equilibria are, therefore, those of a very crudely coupled atmosphere–ocean system (Power and Kleeman 1993). In addition, we show that these equilibria are very sensitive in certain areas. Perturbations in those areas can initiate a transition to another equilibrium or can generate variability on time scales of a decade or longer. It is demonstrated that this behavior is strongly related to the local behavior of the convective process under mixed boundary conditions.

The general outline of the article is as follows. The model is described in section 2. The spinup circulation and the model response to a slowly varying external forcing is studied in section 3. When slowly varying the atmospheric forcing the model response is not smooth, but reveals sudden transitions, associated with sudden changes in the location of deep-water formation. When slowly returning the forcing to original values hysteresis occurs. We were able to find many different equilibria for the same boundary conditions. These equilibria differ with respect to the strength and location of the deep-water formation. The mechanisms governing the deep-water formation process are strongly nonlinear. In section 4 it will be shown that in a one-dimensional box model this nonlinearity gives rise to the occurrence of hysteresis and the possibility of two equilibria, one with and one without convection. In section 5 we use the box model to perform a local stability analysis of the ocean circulation of our three-layer model. It shows that a large area exists in the model ocean where convection does not occur but can be triggered easily. Similarly, a region exists where convection can be suppressed easily. These potentially convective and potentially nonconvective regions contain a simple mechanism for multiple equilibria and variability on decadal time scales.

The box model also provides a tool to draw qualitative conclusions about the significance of some parameters for the behavior of the deep-water formation process. In this respect the time scale of the atmospheric coupling and the salinity and temperature stratification are considered.

2. Model description

The model is based on the geostrophic model, described by Killworth (1985). As shown in Fig. 1, it consists of three layers, labeled with s, 1, and 2, of constant depth h. The basin is a midlatitude β plane centered around 45°N. The longitudinal extension x is L and the latitudinal extension y is M. For each layer the model allows continuously varying temperature (T), salinity (S), and velocity distributions in the hor-

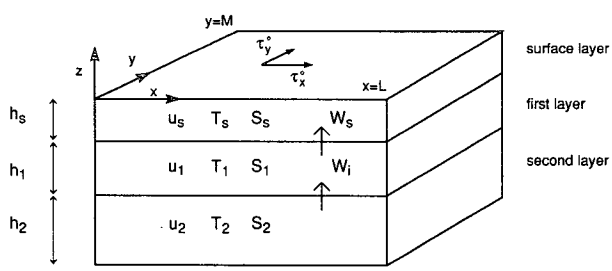


FIG. 1. Ocean model consisting of three layers, denoted by subscripts s, 1, and 2, and a horizontal basin between $x = 0$, L (west to east direction) and $y = 0$, M (south to north direction). Each layer with thickness h has a uniform temperature T, salinity S, and velocity ($\bar{u}$) in the vertical. The vertical velocity is denoted by w. The wind stress τ^0 acts on the surface layer only.

izontal. Within a layer the temperature, salinity, and horizontal velocity fields are assumed to be independent of depth. The vertical velocities are computed at the intersections between the layers.

a. Model equations

The equations for the salinity S and temperature T of a layer are given by the conservation laws for salt and heat, respectively. Here, we give the equations for the surface layer s in flux formulation:

$$\frac{\partial S_s}{\partial t} + \nabla_h \cdot (\bar{\mathbf{u}}_s S_s) - w_s S_s^* / h_s = \kappa_h \nabla_h^2 S_s + D_s^S + F_s^S,$$

$$\frac{\partial T_s}{\partial t} + \nabla_h \cdot (\bar{\mathbf{u}}_s T_s) - w_s T_s^* / h_s = \kappa_h \nabla_h^2 T_s + D_s^T + F_s^T, \tag{1}$$

where ∇_h is the horizontal gradient operator, $\bar{\mathbf{u}}_s = (u_s, v_s)$ is the horizontal velocity vector, and w_s is the vertical velocity at the interface. The terms on the left-hand side are, respectively: the local time derivative, the horizontal flux divergence, and the vertical flux divergence. The terms denoted with the asterix are given by upstream differentiating of the vertical flux divergence, for example,

$$S_s^* = \begin{cases} S_1, & \text{if} \quad w_s > 0 \\ S_s, & \text{if} \quad w_s < 0. \end{cases}$$

On the right-hand side are the horizontal eddy diffusion with constant κ_h, the vertical mixing term D_s, and the atmospheric forcing term F_s, respectively. Here D_s contains vertical diffusion as well as deep-water formation. Vertical diffusion can only poorly be resolved by the used vertical resolution and will probably be underestimated. The upstream scheme, however, is responsible for an additional diffusion with strength $w^*h^*/2$, where w^* is a characteristic vertical velocity and h^* is a characteristic vertical length scale. Typical values of $h^* = 200$ m and $w^* = 10^{-6}$ m s^{-1} yield an extra diffusion with diffusion coefficient 10^{-4} m^2 s^{-1}. The equations for the two other layers are similar.

For the momentum equations we assume a balance between Coriolis force, pressure gradient, and some, generally small, friction terms. For scales as small as 100 km the Rossby number is only of order 0.01. Therefore, it is justified to neglect the local time derivative and the advection terms in the momentum equations. This approximation has been employed by many other authors (e.g., Hasselman 1982; Killworth 1985; Zhang et al. 1993). The friction terms consist of the wind stress, working on the surface layer only, and a Stommel type, linear friction term in the y-momentum equation. This linear friction term with strength κ_{st} is used to permit a western boundary current. The momentum equations for the surface layer are given by

$$-fv_s = -\frac{1}{\rho_0}\frac{\partial p_s}{\partial x} + \frac{\tau_x^0}{h_s \rho_o}$$

$$fu_s = -\frac{1}{\rho_0}\frac{\partial p_s}{\partial y} + \frac{\tau_y^0}{h_s \rho_o} - \kappa_{st} v_s. \tag{2}$$

Here $\bar{\tau}^0$ is the wind stress vector, which acts on the surface layer only; p_s is the pressure; and $f = f_0 + \beta y$ is the Coriolis parameter, where f_0 and $\beta = df/dy$ are evaluated at 45°N. The vertical velocities are computed from the continuity equation.

The equation of state used is a fit, linear in S and quadratic in T, to the UNESCO formula (1981) and is designed to represent the density gradients well. A fit of the same form with similar coefficients has been designed previously by Bryan and Cox (1972). The pressure dependency has been ignored:

$$\rho = \rho_0 + k_S S - k_T (T - T^*)^2. \tag{3}$$

The constants are, respectively: $k_S = 0.7739$ kg m^{-3} per promille, $\kappa_T = 0.00471$ kg m^{-3}K^{-2}, $T^* = 265.42$ K, and $\rho_0 = 1001.3263$ kg m^{-3}. The units of T and S are kelvin and psu (promille), respectively.

b. Method of solution

The horizontal velocity is split into a barotropic and two baroclinic modes. The barotropic mode is completely determined by the wind stress. The baroclinic modes are computed from the density distribution. The diagnostically determined velocities are used in the time evolution of the temperature and salt fields. For the time discretization of the tracers T and S a leapfrog method is used for the contribution of the horizontal advective terms and an Eulerian scheme for the contribution of the diffusive terms. Since, by the geostrophic assumption, the fast waves have been filtered out, a relatively large time step of approximately five days can be used.

The horizontal discretization is done on an Arakawa C grid. Special care has been taken so that the continuity equation is exactly satisfied on this grid. The scheme also conserves S and T exactly.

The process of deep-water formation is parameterized by a convective adjustment procedure applied after each time step. Starting from the uppermost two layers it is checked if they are unstably stratified, and in that case these layers are completely mixed for T and S. Next, the same procedure is applied to the lowest two layers. This procedure does not guarantee complete stability of the whole column. Recent studies, however, indicate that this does not seriously affect the stability properties of the thermohaline circulation (Weaver et al. 1993). This convective procedure has an associated time scale of approximately twice the time step of the ocean model, thus being a (few) week(s). Although this time scale might seem too long, considering the convective time scale on a small (chimney) length scale,

it seems defendable on the larger (gridbox) length scale. In this way the time scale of the processes that homogenize the fluid over larger scales (Killworth 1983) may be taken into account.

c. Forcing

Crucial in models of the thermohaline circulation is the use of mixed boundary conditions consisting of a relaxation condition for the surface temperature and a prescribed salt flux for the surface salinity. The freshwater flux resulting from the net effect of precipitation and evaporation and river runoff is translated into an effective salt flux.

To compute the freshwater flux, ocean modelers usually spin their models up under restoring conditions for both temperature and salinity. The freshwater flux in equilibrium is diagnosed and for the rest of the integration kept constant. It is argued that this procedure compensates for missing physics in the model and it produces a value for the freshwater flux where measurements are not available or insufficient. An additional argument is that the model converges much slower toward equilibrium for mixed boundary conditions than for relaxation boundary conditions. For large models this time aspect makes a spinup under mixed boundary conditions hardly possible. By using relaxation conditions for both salinity and temperature, the salinity fluxes in equilibrium strongly depend on the model formulation. With a high-resolution, realistic model this procedure may lead to a realistic precipitation minus evaporation field. With a simple low-resolution model, however, one can never expect realistic freshwater fluxes from this procedure. We therefore used mixed boundary conditions during the spinup. This method makes it possible to use freshwater fluxes with a relatively simple spatial distribution, whereas those obtained under restoring conditions usually have a very complex spatial distribution. Considering the known sensitivity of the model response to details of the freshwater flux (e.g., Weaver and Sarachik 1991; Weaver et al. 1993), the complex distribution makes it difficult to evaluate and compare model results. Prescribing a freshwater flux of a relatively simple form makes comparison easier and makes it possible to investigate the influence of the magnitude and the spatial distribution of the salt forcing more precisely.

For the salinity flux a simple cosine profile is used, representing net evaporation near the equator and net precipitation near the pole:

$$F_s^S = A_S \cos \frac{\pi y}{M}, \tag{4}$$

where A_S is a constant, chosen such that the magnitude of the freshwater flux is in the order of 100 mm mo^{-1}.

The temperature of the surface layer is relaxed with a rate α toward an apparent atmospheric temperature following a cosine law in latitude:

$$F_s^T = \alpha(T_{\text{atm}} - T_s),$$

with

$$T_{\text{atm}} = T_m(0.5 + 0.5 \cos \frac{\pi y}{M}). \tag{5}$$

The zonal wind stress is a function of latitude only and roughly represents the observed climatological mean meridional distribution of the zonal winds:

$$\tau_x = -\frac{\rho_0 f M}{2\pi} w_{s0} \cos \frac{2\pi y}{M}, \tag{6}$$

where w_{s0} is a constant. The meridional wind stress is neglected.

d. Discretization

The horizontal dimensions of the basin we have considered are 5000×5000 km^2. The horizontal grid consists of 20×20 points. The thicknesses of the layers are $h_s = 50$ m, $h_1 = 400$ m, and $h_2 = 4000$ m. For the vertical diffusion coefficient κ_v and horizontal diffusion coefficient κ_h we chose the commonly used values 1×10^{-4} m^2 s^{-1} and 3×10^3 m^2 s^{-1}, respectively. The Stommel friction coefficient κ_{st} is 4×10^{-6} s^{-1}, chosen such that the boundary current is just resolved on the used grid.

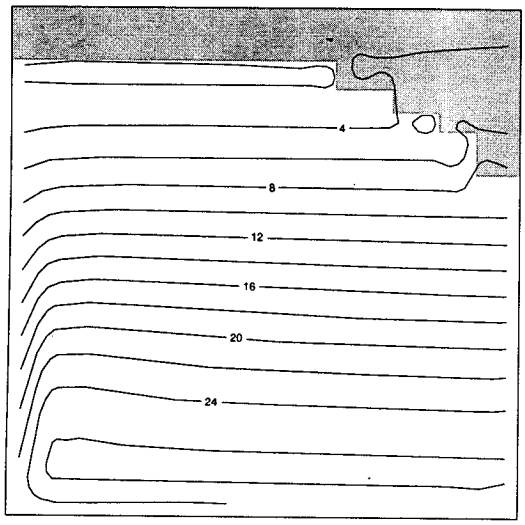
3. Experiments

a. Spinup circulation

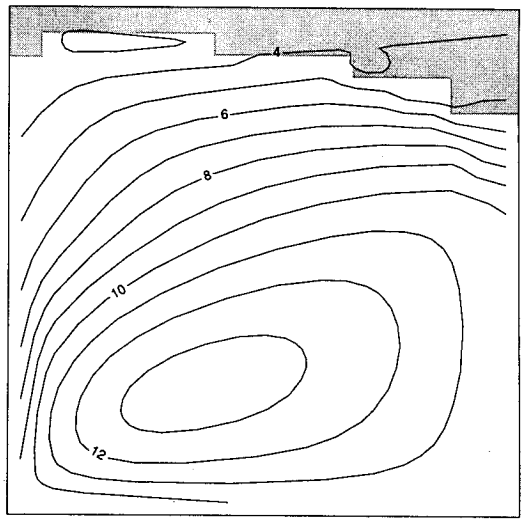
Starting from uniform initial conditions the model was run under mixed boundary conditions until it reached a quasi-stationary state. For the forcing fields we used

$T_m = 28$°C, yielding a pole to equator temperature gradient of 28°C, $\alpha = (50 \text{ days})^{-1}$,
$A_s = 3 \times 10^{-8}$ psu s^{-1} corresponding to about 110 mm mo^{-1} net precipitation, and
$w_{s0} = 5 \times 10^{-7}$ m s^{-1}, which is equivalent to a surface wind maximum of about 10 m s^{-1}.

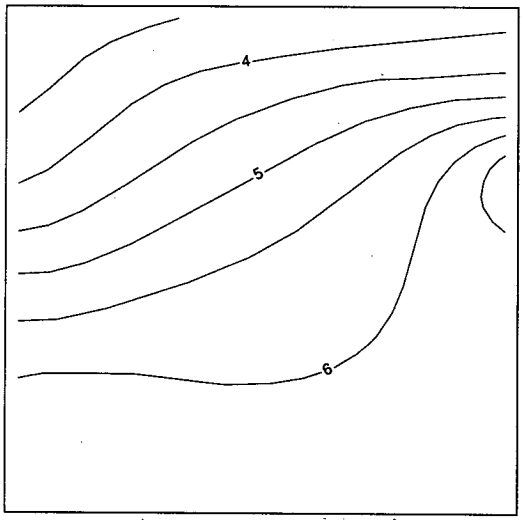
After a transient period of approximately 1000 years, our model reached a quasi-stationary state with a net heat uptake by the ocean of less than 5×10^{-3} W m^{-2}. The temperature and horizontal velocity fields of this equilibrium, referred to as E_1, are shown in Fig. 2. Although the circulation has unrealistic features, such as the relatively strong vertical velocities near the eastern and western boundary, the overall circulation looks reasonable. In the surface layer a broad boundary current flows along the western boundary until halfway poleward of the basin. There it starts to cross the ocean northeastward. About half of this current flows back southward in a subtropical gyre. Deep water is formed in the north and northeastern part of the basin. In this deep-water formation region the oceanic heat loss is



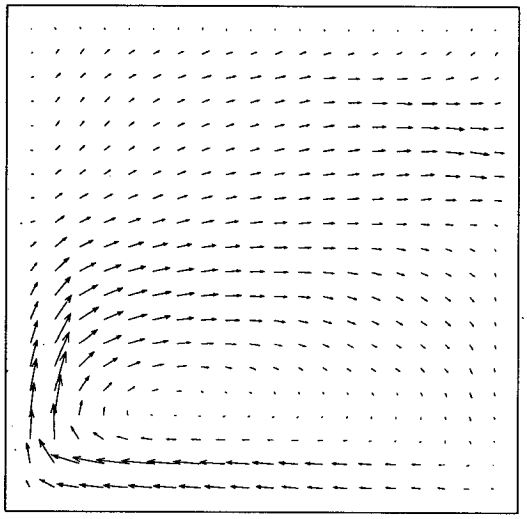
temperature ekman layer



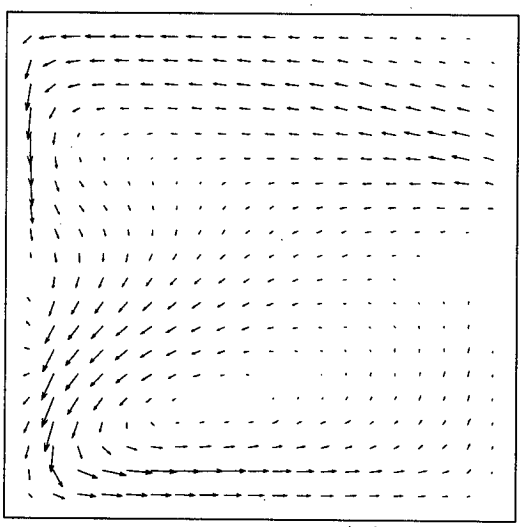
temperature first layer



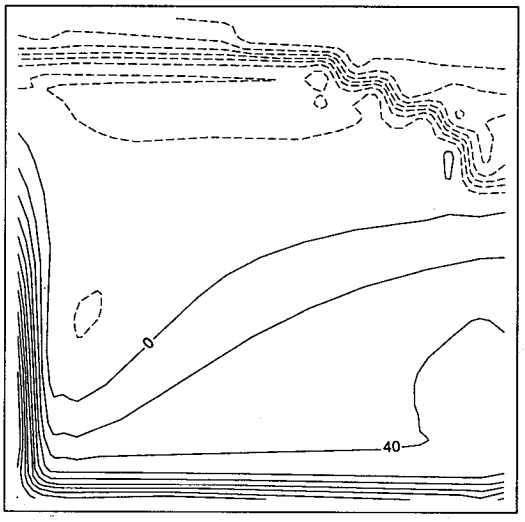
temperature second layer



velocity ekman layer



velocity second layer



relaxation

large. The properties of the deep ocean layer are mainly determined by the deep-water formation process. A western intensification can be observed in the temperature, salinity, and velocity fields. The western and southern boundary regions of the surface layer remain relatively cool and fresh (low salinity), due to the high vertical velocity, which advects cool and fresh deep ocean water.

b. *Polar halocline catastrophe*

In previous studies (e.g., Bryan 1986a,b; Marotzke 1990) it has been shown that the spinup circulation is extremely sensitive to salt perturbations in the polar region after switching over to mixed boundary conditions. To test whether or not this holds for our simple model, which has been spun up under mixed boundary conditions using a spatially simple freshwater flux, we applied a salt anomaly of −0.2 psu in the northern thousand kilometers of the surface layer. Similar as described by Bryan (1986a,b) a polar halocline catastrophe occurs. Here, the description is deliberately kept short because the basic physical mechanisms are relatively simple and have been extensively described by the referred authors. The applied anomaly is large enough to completely suppress convection. Since the deep-water formation is the main source of density differences, the meridional overturning now collapses in about ten years and a weak reversed circulation settles. Diffusive processes cause a steady warming of the deep ocean layer. The basin mean heat absorption rate during this period is approximately 10 W m^{-2}. After about 140 years the stratification in the subpolar regions becomes unstable and convection sets in. The convective process intensifies now rapidly, and as a result a strong positive circulation emerges. During this so-called flush, the ocean quickly loses heat with an initial rate of about 40 W m^{-2}. Hereafter, the ocean circulation slowly converges to a state that largely resembles the spinup state. The entire process is shown in Fig. 3 for the meridional heat transport and the basin mean temperature of the ocean.

c. *Time-varying forcing*

The first part of this experiment consisted in slowly reducing the meridional temperature gradient T_m from 28° to 23°C in a time span of 1500 years. As initial condition we used the spinup circulation. During the time integration the ocean circulation revealed sudden transitions. The meridional heat transport for instance, as shown in Fig. 4, displays a decrease of about 10%

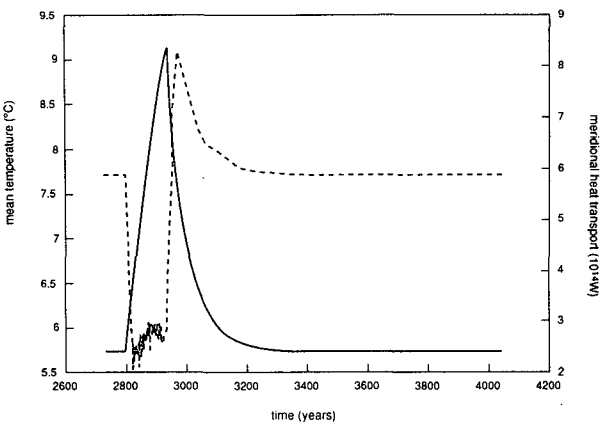


FIG. 3. Integrated meridional heat transport at $y = M/2$ in 10^{14} W m^{-2} (dashed line) and basin mean temperature (°C) (solid line) during polar catastrophe.

around 27.5°C. Smaller changes occur around 24.2° and 23.8°C. These transitions are characterized by sudden changes between convective regimes. The area where convection occurs changes dramatically during the transitions, whereas it remains almost constant during the rest of the integration period. Typical changes in the convective area during a transition vary from 3×10^5 to 12×10^5 km^2 (5–20 grid boxes). The final circulation, resulting from the previous time integration, was run under constant forcing (T_m = 23°C) to a stationary state with a net heat uptake of the ocean of less than 5×10^{-3} W m^{-2}. The changes in circulation during this time integration were only small, thus indicating that the rate of change of the forcing was slow enough to permit a quasi-stationary behavior during the variation experiment. This equilibrium was used as an initial condition for a time integration in which the temperature gradient was increased during again 1500 years from 23° to 28°C. The results are shown by the dashed line in Fig. 4. The transitions occurred for different values of T_m compared with the previous time integration. The final equilibrium for T_m = 28°C, which will be denoted by E_2, was also run to a stationary state with a total heat loss of the ocean of less than 5×10^{-3} W m^{-2}. As shown in Fig. 5 this steady state is fairly different from the spinup steady state, E_1, obtained for the same restoring temperature. The two equilibria thus obtained differ most predominantly in the regions of deep-water formation. In E_1 deep water is formed in a broad band near the polar boundary, slightly going southward when approaching the eastern

FIG. 2. Horizontal plane with temperature distribution of the surface (Ekman) (a), first (b), and second layer (c) [in the shaded regions convective mixing takes place between the surface and the first layer (a) and between the first and the second layer (b)]; the horizontal velocity distribution of the surface (d) (maximum velocity 0.12 m s^{-1}) and the second layer (e) (maximum velocity 0.01 m s^{-1}); and the atmospheric heat flux (f) in watts per square meter for equilibrium E_1.

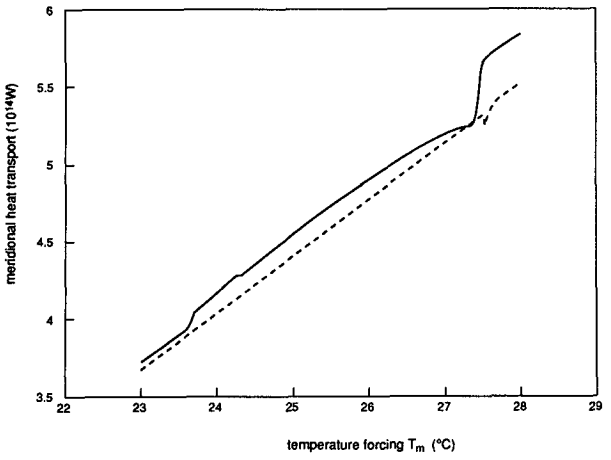


FIG. 4. Integrated meridional heat transport at $y = M/2$ during the variation experiment. The solid line is the integration from 28° to 23°C; the dashed line is the backward integration.

boundary. State E_2 has bottom water formation in a small band near the polar boundary and in a large area in the northeastern part of the basin. The different rate and location of deep-water formation are responsible for the different temperature of the deep ocean; the deep ocean of E_2 is about 1.0 degree warmer than E_1. Deep-water formation is a major source of density and thus pressure differences. Consequently, there are differences between the velocity distribution of E_1 and E_2. These differences are, however, mainly distinguishable in the polar regions of the basin. Since halfway into the basin the meridional velocity is not significantly affected by the differences in convective regime, the differences in meridional heat transport ($\overline{vT}$) are mainly caused by temperature differences. The surface temperature displays only small differences between the two states due to the strong relaxation toward the atmospheric temperature. Henceforth, differences in meridional heat transport are mainly caused by differences in the lower-layer temperature. As a consequence of the 1.0 degree warmer bottom temperature of E_2, the ocean circulation transports about 10% less heat northward as compared to E_1.

A similar experiment was performed starting from E_2. After varying the forcing again backward and forward, we obtained another equilibrium. Starting from this equilibrium the whole cycle was again repeated and again a new equilibrium was resulting. In this way we were able to find many different equilibria. During this iterative process we did not touch upon an equilibrium previously found. The equilibria found are sometimes rather similar. Nevertheless, as shown in the first experiment, substantial differences can occur.

4. Multiple equilibria in convective activity

The sudden transitions between the equilibria are characterized by changes in convective activity. As deep-water formation seems to be the dominant process, we will now focus on this local process using a simple box model. The model is similar to box models designed by Welander (1982) and Marotzke (1990).

a. Box model

As shown in Fig. 6, the model consists of a surface layer box with varying temperature T and salinity S above a bottom box b with a constant temperature T_b and salinity S_b. The horizontal exchange of salt and heat between the surface-layer box and its neighbor i is parameterized as

$$q(S_i - S)$$

and

$$q(T_i - T),$$

where q^{-1} is the time scale of this process, which is determined by the advective and diffusive time scales, and S_i and T_i are constants. In analogy with the ocean model, mixed boundary conditions are prescribed. The temperature T is relaxed toward the atmospheric temperature T_{atm} with a time constant α^{-1}, which is in the order of a month. A constant salinity flux F^S is prescribed for the surface salinity. Exchange with the bottom box b occurs only if the surface box is denser than the bottom box. The exchange is in this case

$$\tau(S_b - S)$$

and

$$\tau(T_b - T),$$

where τ^{-1} is the time scale of the convective adjustment, which is in the order of a few days to weeks. This model yields the following set of equations:

$$\frac{dS}{dt} = F^S + q(S_i - S) + \tau\Delta(\delta\rho)(S_b - S), \quad (7)$$

$$\frac{dT}{dt} = \alpha(T_{\text{atm}} - T) + q(T_i - T) + \tau\Delta(\delta\rho)(T_b - T), \quad (8)$$

in which $\delta\rho = \rho - \rho_b$ is the density difference between surface box and bottom box b. Similar to the ocean model Δ is a stepfunction:

$$\Delta(\delta\rho) = 1, \quad \text{if} \quad \delta\rho > 0$$
$$= 0, \quad \text{if} \quad \delta\rho < 0.$$

Vertical advection is not considered in the analysis. It can, however, easily be included without essentially affecting the results.

b. Equilibrium solutions of the box model

It can easily be shown by linear stability analysis that the stationary points of equations (7) and (8) will

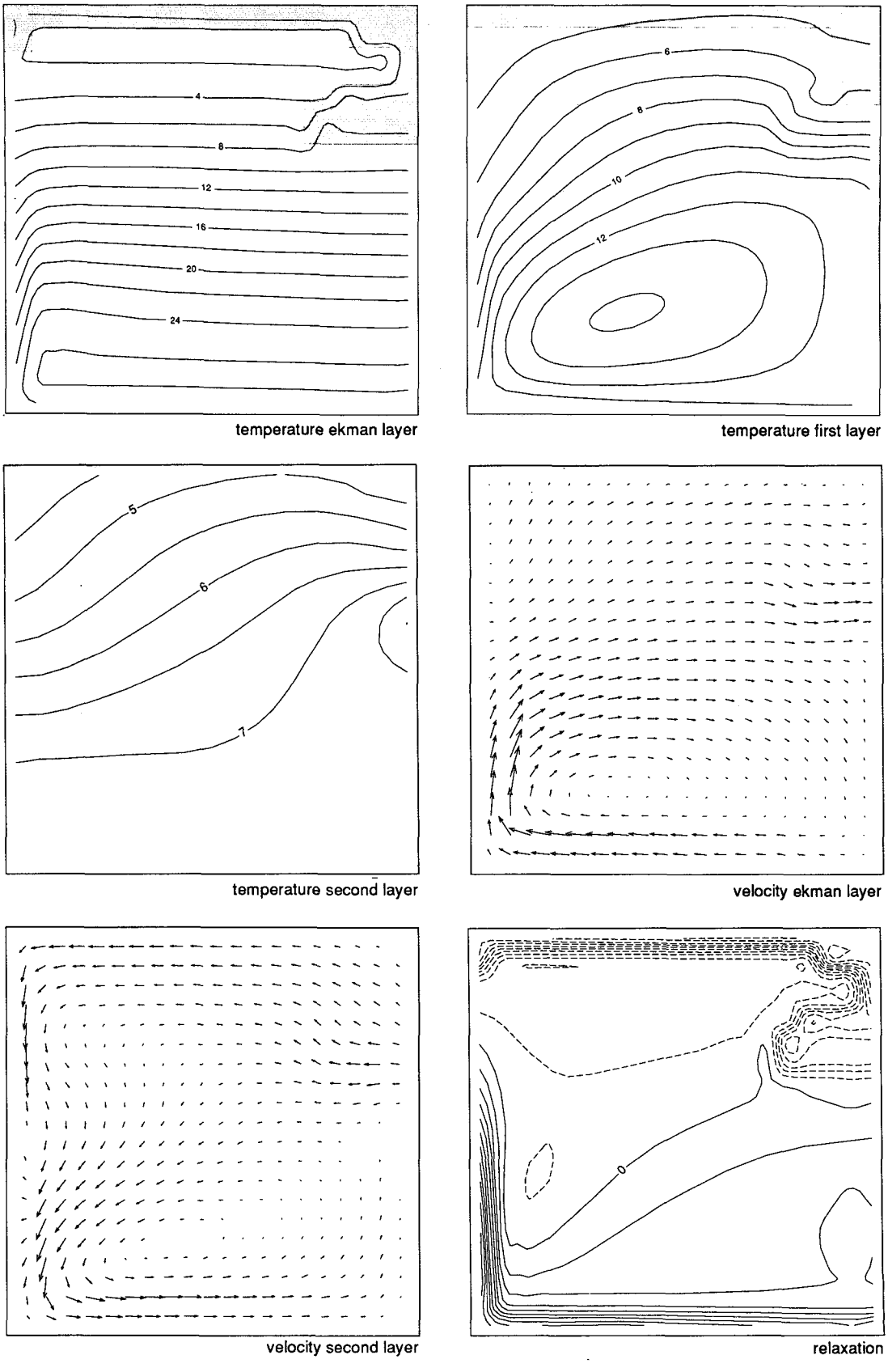


FIG. 5. As Fig. 2 but for equilibrium E_2.

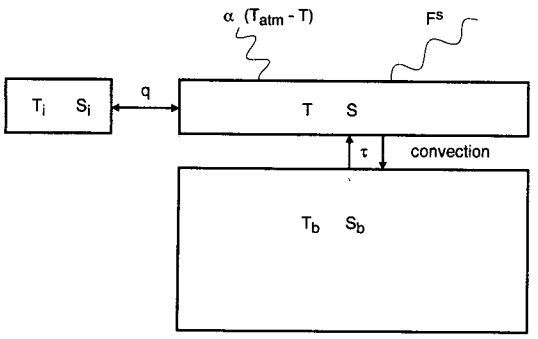


FIG. 6. Surface box with time-varying T and S, surrounded by a box with constant temperature T_i and salinity S_i and transfer coefficient q. Convective exchange with time constant τ^{-1} occurs if the surface water becomes denser than the deep water. The atmospheric forcing consists of a relaxating flux toward T_{atm} and a prescribed salinity flux F^S.

be stable. The stationary points can be obtained by setting the left-hand sides of (7) and (8) to zero. Thus, we obtain

$$T_{\text{eq}}(\Delta) = \frac{qT_i + \alpha T_{\text{atm}} + \tau\Delta T_b}{q + \alpha + \tau\Delta}, \tag{9}$$

$$S_{\text{eq}}(\Delta) = \frac{qS_i + F^S + \tau\Delta S_b}{q + \tau\Delta}. \tag{10}$$

The convective equilibrium exists if and only if

$$\rho(T_{\text{eq}}(1), S_{\text{eq}}(1)) > \rho_b, \tag{11}$$

whereas the nonconvective equilibrium exists if and only if

$$\rho(T_{\text{eq}}(0), S_{\text{eq}}(0)) < \rho_b. \tag{12}$$

As the range over which the temperature varies in the convective region is only small, we use a linear equation of state to facilitate the derivation:

$$\rho(T, S) = \rho_0 - k_T^* T + k_s S,$$

where ρ_0, k_T^*, and k_S are positive constants. The condition for the existence of the convective equilibrium becomes

$$\frac{k_S[F^S + q(S_i - S_b)]}{q + \tau} - \frac{k_T^*[\alpha(T_{\text{atm}} - T_b) + q(T_i - T_b)]}{q + \alpha + \tau} > 0,$$

which states that the density stratification caused by the salinity fluxes plus the density stratification caused by the temperature fluxes must be positive. Analogously, the condition for the existence of the nonconvective equilibrium becomes

$$\frac{k_S[F^S + q(S_i - S_b)]}{q} - \frac{k_T^*[\alpha(T_{\text{atm}} - T_b) + q(T - T_b)]}{q + \alpha} < 0.$$

For the convective equilibrium we rewrite this as

$$\Phi^S > -\kappa(\tau)\Phi^T,$$

and for the nonconvective equilibrium:

$$\Phi^S < -\kappa(0)\Phi^T,$$

where the following abbreviations were made:

$$\Phi^S \equiv k_S[F^S + q(S_i - S_b)]$$

$$\Phi^T \equiv -k_T^*[\alpha(T_{\text{atm}} - T_b) + q(T_i - T_b)]$$

$$\kappa(\tau) \equiv \frac{q + \tau}{q + \alpha + \tau}.$$

For future reference we remark that

$$\kappa(\tau) - \kappa(0) = \frac{\alpha\tau}{(q + \alpha + \tau)(q + \alpha)} > 0.$$

From these equations we can determine the equilibrium solutions of the box model. Schematically, these are shown in Fig. 7. Dependent on the sign of Φ^T, we can distinguish two different cases. If Φ^T is negative then three regimes can be distinguished. For Φ^S sufficiently large [$\Phi^S > -\kappa(\tau)\Phi^T$] only convection is possible (regime I), whereas for sufficiently small values

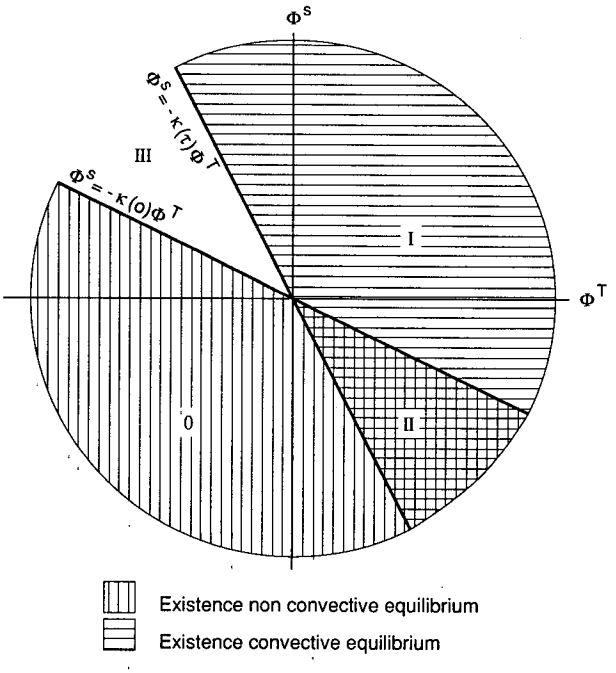


FIG. 7. Survey of the possible equilibria of the box model.

$\Phi^S < -\kappa(0)\Phi^T$ only the nonconvective equilibrium is possible (regime O). In between is regime III, where no equilibria exist. In this regime oscillations occur between the convective and the nonconvective state. The heat salt oscillator described by Welander (1982) is an example of this regime. Similarly, when Φ^T is positive, we can distinguish three regimes. The convective regime I exists for $\Phi^S > -\kappa(0)\Phi^T$, whereas for $\Phi^S < -\kappa(\tau)\Phi^T$ the nonconvective regime O exists. For $-\kappa(\tau)\Phi^T < \Phi^S < -\kappa(0)\Phi^T$ both the convective and nonconvective equilibria exist. This regime will be denoted with II.

In physical terms the existence of the regime II can easily be explained. Consider an initial state, which is only marginally stable and a stratification of relatively cold and freshwater above warmer and saltier water. Furthermore, we assume that the atmosphere has a cooling effect and that a positive freshwater flux acts on the surface water. For the polar regions of the ocean this is the usual situation. If somehow convection is triggered, then cold and freshwater is mixed downward and at the same time the surface water becomes saltier and warmer. Due to the strong relaxation toward T_{atm} the surface layer rapidly loses the additional heat and thus becomes denser, whereas the surface salinity remains higher than before the onset of convection. This yields a positive feedback giving rise to sustained convection. This mechanism or similar ones have also been recognized by several authors (Power 1994; Zhang et al. 1993; Wright and Stocker 1991).

A practical consequence of regime II is the occurrence of hysteresis. This hysteresis will be called *local* hysteresis of the convective process. To illustrate this, we have chosen to vary Φ^S slowly from a value somewhat smaller than $-\kappa(\tau)\Phi^T$ to a value somewhat larger than $-\kappa(0)\Phi^T$ and backward again. For $\Phi^S < -\kappa(\tau)\Phi^T$ only the nonconvective equilibrium exists and convective mixing does not occur. This situation remains stable until Φ^S becomes larger than $-\kappa(0)\Phi^T$. Here, only the convective equilibrium exists and the system will inevitably turn into the mixing state. Convective mixing continues until Φ^S is again smaller than $-\kappa(\tau)\Phi^T$. The resulting local hysteresis of the convective process is shown in Fig. 8.

The breadth of regime II is $(\kappa(\tau) - \kappa(0))|\Phi^T|$, which is exactly equal to $k_T^* \alpha(T_{\text{eq}}(1) - T_{\text{eq}}(0))$. This expression can be considered as the difference between the density gain by the atmospheric forcing in the convective state and density gain in the nonconvective state, and will be called the potential density gain. In the polar region of the model ocean, the atmospheric forcing is in the convective region order 100 W m^{-2} and in the nonconvective regions order 10 W m^{-2} (Fig. 2f). So, the potential density gain by convection is large, which yields a broad regime II.

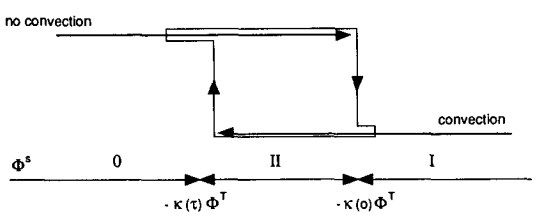


FIG. 8. Hysteresis loop of the box model.

5. Stability analysis of the ocean model using the box model

a. *Identification of the regimes in the ocean model*

We have identified the different regimes that are possible in the steady states of the ocean model. For each gridpoint we performed the following procedure. Assuming constant salinity and temperature for the surrounding gridpoints and constant transfer coefficients, we applied the box model and determined in which of the regimes, discussed in the previous section, that particular gridpoint is. Here, we drop the assumption of one surrounding box, which makes the analysis of the previous section somewhat more extended but does not change its basics. The transfer coefficients q are determined by the flow and the diffusion coefficients.

The results for equilibrium E_1 are shown in Fig. 9. In this figure we observe a large region satisfying the conditions of regime II. A small band of this region is situated at the border of the region where convection takes place. This area is being defined as the potentially nonconvective area. Our analysis indicates that, besides the convective state, the nonconvective state is also a local solution (as long as the surrounding area does not change too much). Small perturbations in the atmospheric forcing or an increase in the advection of lighter water can suppress convection in the potentially nonconvective area and cause the convective border to draw back northward. The remainder of the convective area satisfies the conditions of regime I, where convection is the only possible solution. Most of the points in regime II are situated outside the area where convection is actually occurring. These points indicate a region of potential convection where convective mixing is possible according to our analysis but not being realized. If convective mixing is triggered in this region it tends to continue. This so-called potentially convective area is located in the northeastern part because the Gulf Stream Extension is transporting relatively saline water into this region; in the northwestern corner this salt advection is absent and the freshwater flux is more dominant. The remaining nonconvective area fullfills the regime O conditions. It should be noted that none of the gridpoints satisfy the conditions of regime III.

b. *Practical consequences of the regimes for the ocean model*

Next, we tested whether the previous analysis has practical consequences for the behavior of the ocean

model. We imposed a small positive salinity perturbation on the steady state E_1 in a single gridpoint (approximately 6×10^4 km^2 and 0.1 psu) in the surface layer and investigated the changes in the ocean circulation during the subsequent time integration. We did this for all the grid points outside the convective region. The salinity anomaly was just large enough to initiate convective overturning. As we expected from our analysis, for points outside the potentially convective area, the initiated convection disappears within a month. In the region of potential convection the time-dependent behavior can be grouped into different classes. For a few gridpoints, located near the border of the potentially convective area, convection continues at that gridpoint only for at least five years. For a perturbation in the other gridpoints the initiated convective area starts to expand after a few months. We will describe the time evolution for a positive salt anomaly imposed in two different gridpoints, indicated by 1 and 2 in Fig. 9. They represent two different types of behavior that are observed. For a perturbation that starts in position 1, convection continues for a period of about a few months. Hereafter the convective zone starts to spread eastward and northward and fills up the whole area eastward and northward of location 1 in a few years. The increased convective activity results in a stronger meridional circulation. As a result of the larger horizontal advection of lighter water, the convective area draws back until its extension is slightly smaller than the initial convective region. The time scale of this feedback process of the circulation is a few years. For an anomaly in position 2 convection also spreads eastward. The feedback of the circulation is much smaller now. Consequently, the new convective area remains stable. This experiment shows that new equilibria can easily be induced by small perturbations at the right location. The results indicate that there are many ways for the convective mixing to adjust in the northeastern region. In a similar way Rahmstorf (personal communication) found different equilibria in a more realistic (high vertical resolution, idealized global basin) ocean model.

The process leading to the spreading of the convective area can be understood as follows. The initiated convection mixes relatively warm and saline water upward. Due to the strong atmospheric relaxation, the surface waters are rapidly cooled. However, such a fast process to remove the salinity anomaly does not exist. Due to the horizontal advection and diffusion, the salinity anomaly starts to extend, favoring the convective mixing in the neighboring areas.

We also investigated the sensitivity of the potentially nonconvective area, that is, the area where deep water is presently being formed but where the nonconvective equilibrium is also possible according to our analysis. As the stratification in this region is only marginally unstable, it is easy to suppress the deep-water formation process by a small negative salt anomaly. In the potentially nonconvective region we perturbated the solution with small salt anomalies in a single grid cell. The anomalies were chosen to be just large enough to suppress convection, which is typically -0.01 psu. Analogous to the previous experiment, where convection tended to spread after being triggered, it turned out that for all gridpoints the initiated nonconvective area has the tendency to expand. However, this tendency is in the northwestern convective region much stronger than in the northeastern region. The western convective area appears to be very sensitive.

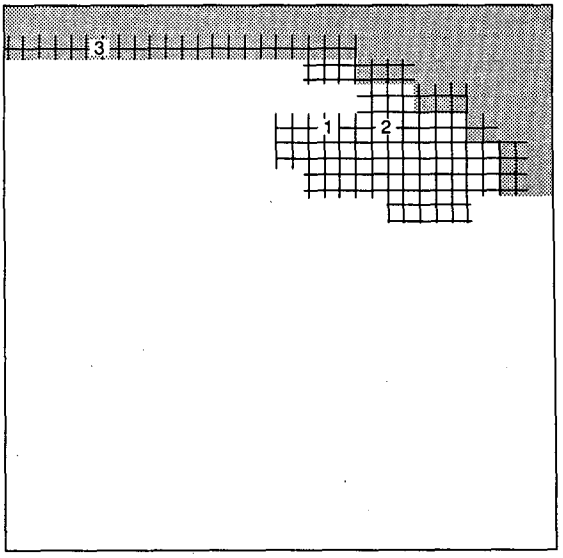


FIG. 9. Results of the box model analysis applied to E_1. Shown are the convective area (shaded) and the area satisfying the conditions of regime II (grid).

c. *Explanation convective transitions*

The results of the perturbation analysis are used to explain the sudden transitions as observed in the variation experiment (section 3c). We will concentrate on the significant transition that occurred between 27.0° and 28.0°C in this experiment. We have again integrated the model, but now the temperature gradient T_m is more slowly being reduced from 28.0° to 27.0°C in 2500 years. Near 27.8°C a transition took place; deep-water formation disappeared for a 1.2×10^6 km^2 large area at the southern boundary of the convective region. The area was mainly located in the northwestern part of the basin. Due to the decreased convective activity the temperature of the bottom water rose about 0.7°C. Consequently the meridional energy transport dropped about 10%, as shown in Fig. 10. The overshoot just after the transition is transient behavior, which lasts about 300 years. This represents the diffusive time scale of the deep ocean to adjust to the new convective equilibrium. Slowly increasing the temperature gradient again to 28°C did not change the area of deep-

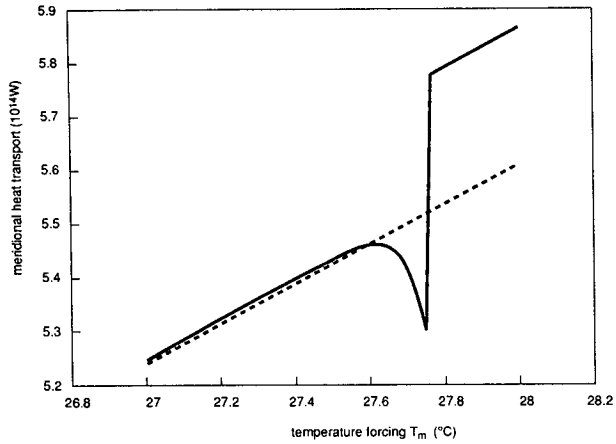


FIG. 10. Integrated meridional heat transport at $y = M/2$. The solid line is the time integration from $T_m = 28°$ to 27°C; the dashed line is the backward time integration.

water formation. The newly found equilibrium at 28.0°C will be denoted shortly by E_3.

The results of the foregoing perturbation experiments showed that deep-water formation in the western region is very sensitive to negative salt anomalies. The nonconvective area, once triggered, extends strongly. We determined the location where convection for the first time during the integration is suppressed. In Fig. 9 it is shown that this location, labeled with a 3, is located in the sensitive area. As predicted by the perturbation experiments, the suppression of convection spreads eastward and to a lesser degree westward in a few years time.

The results of the box analysis of the new equilibrium E_3 are shown in Fig. 11. A small deep-water formation area is surrounded by a large potentially convective area. The large size of the potentially convective area is mainly due to the high bottom temperature of E_3. A high bottom temperature implies a large potential density gain by convection and thus a broad local hysteresis. Spatially a large potentially convective area results. Perturbation experiments, similar to those previously done, showed that the potentially convective area in the northeastern corner is sensitive in the sense that the triggered convective area strongly tends to expand. In the northwestern corner the triggered convective area remains locally bounded.

The view that emerges to explain the difference in the shape of the deep-water formation region of the spinup circulation E_1 and that of E_2 is as follows: the spinup convective region is only marginally stable in the northwestern corner. Convection disappears here when the heat forcing is reduced. Increasing the forcing again leads to a buildup of a convective area in the northeastern corner, whereas a corresponding increase in the northwestern corner does not take place.

6. Discussion and conclusions

In this paper we showed that the thermohaline circulation of the ocean model displays multiple steady states within the thermally driven regime. These multiple steady states were obtained by a slow variation of the heat forcing. During certain stages of this process the ocean circulation reveals sudden jumps. These jumps are strongly related to the deep-water formation process. By slowly varying the heat forcing backward and forward, hysteresis between different steady states was uncovered. Such hysteresis is not unusual and has for instance been discussed by Stocker and Wright (1991).

The sudden transitions are caused by the existence of regime II in the local process of deep-water formation. For the ocean this results in a potentially convective region, in which convection is possible but not occurring. If, by any cause, convection is triggered in this region, it will continue without requiring a feedback from the circulation. Similarly, there also exists a potentially nonconvective region, where convection can be suppressed easily. Within these regions there are a few areas where the on/off switching of convection remains a local process and does not change the state of convection in the neighboring areas. However, for most of the area satisfying the conditions of regime II, a "domino" effect is possible. Triggering or suppressing of convection in a small area results in the change of convective state in a much larger region. This occurs during the sudden transitions in the ocean circulation when the heat forcing is slowly changed.

The sudden change in the area where convection actually takes place will have its effect on many aspects

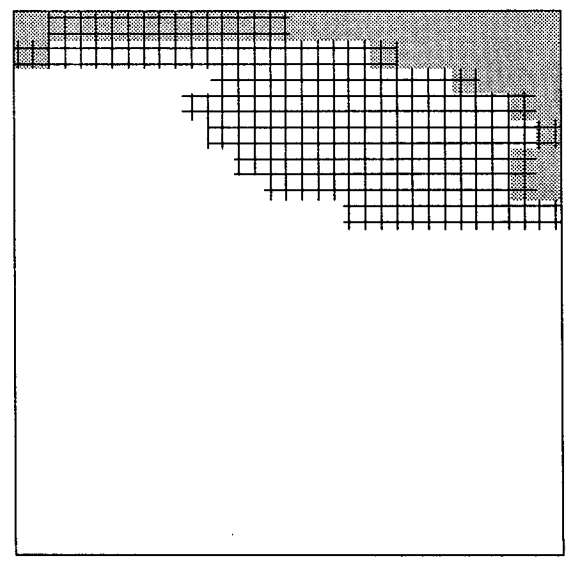


FIG. 11. Results of the box model analysis applied to E_3. Shown are the convective area (shaded) and the area satisfying the conditions of regime II (grid).

of the ocean circulation. Here, we mention two important aspects. The deep-ocean temperature and, simultaneously, the meridional energy transport are mainly determined by the amount and temperature of the deep water formed in the convective area. The process of deep-water formation is also a major source of density differences. Consequently, by the geostrophic relationship, the velocity distributions have to adjust to the new convective equilibrium. The occurrence of the polar halocline catastrophe, for instance, shows the relevance of both aspects.

The relevance of the change in convective area for the variability of the ocean can be estimated by the spatial extension of the area where convection is affected times its local influence. As a measure for the local influence the breadth of the local hysteresis cycle seems to be an appropriate choice. The local potential density gain by convection, which is defined as the difference in density gain by the atmospheric heat flux between the convective and the nonconvective state, is equal to the breadth of the local hysteresis. Neglecting the nonlinear effects of the equation of state, the potential density gain is linear with the rate of density change of the whole column. It should carefully be noted that the convective process, in itself, does not change the average density of the column; it only redistributes density. Delworth et al. (1993) argued that the average density in the sinking region is a forerunner of the strength of the thermohaline circulation. Therefore, the potential density gain and thus the local hysteris breadth seems to be a useful quantity in estimating the influence of convection on the thermohaline circulation. The local hysteresis breadth is strongly dependent on the convective time scale τ^{-1} and the atmospheric relaxation time scale α^{-1}. The extension of the area satisfying regime II, which is primarily the area where these sudden changes can occur, is also dependent on these time scales. Decreasing α does not only result in a decrease in local potential density gain/loss but also decreases the area in which the convective cascade can occur. This will result in a decrease in sensitivity of the thermohaline circulation. This agrees well with the results of Power and Kleeman (1994) and Zhang et al. (1993), who showed that the polar halocline catastrophe became less pronounced when α was decreased. In the extreme case, when a heat flux is being prescribed, regime II does not exist. The potential density gain by convection is zero. Under these boundary conditions the ocean circulation is almost perfectly stable. Based on our analysis we also expect the convective time scale τ^{-1} to have a crucial influence on the sensitivity of the thermohaline circulation.

The behavior of the ocean circulation was strongly related to the existence of the local hysteresis of the convective process. This local hysteresis only exists under certain conditions. Roughly speaking, these conditions assume a stratification of cold and freshwater above saltier and warmer water and cooling and net precipitation by the atmosphere. On the other hand, if the sign of the temperature and salinity stratification and of the atmospheric forcing is reversed, then the local convective process may exhibit periodic behavior (see Fig. 7, regime III). Due to the nonlinearity in advection, the ocean model may respond on various time scales. These remarks may seem obvious; the impacts, however, can easily be overlooked. Consider, as a simple test case, a circulation run under mixed boundary conditions using a very small (10% of the original value) freshwater flux. As a consequence the salinity stratification is nearly zero; the temperature stratification in the convective area is also small. Here, it should be kept in mind that in the convective case the temperature and salinity stratification are coupled by the fact that the density stratification is (nearly) neutral. A relatively large area of deep-water formation with a relatively small heat loss per unit area is resulting. Applying the box model analysis, we found that the area satisfying the conditions of regime II is very small. The ocean circulation proved to be stable; a polar halocline catastrophe could not be induced.

Given the importance of deep-water formation for the maintenance of the thermohaline circulation, one may conclude that variability and stability of the thermohaline circulation is strongly linked to the salt and temperature stratification in the polar regions. The salinity budget of the polar ocean is strongly influenced by the forming of sea ice. During winter the density of the polar water not only increases by cooling but also by salt rejection during sea ice formation, which will thus influence the stratification.

The magnitude of the processes associated with the convective overturning in this model are probably overestimated. However, from measurements, it is known that the heat loss of the ocean is much larger in convective areas than in regions with no overturning activity, a fact that agrees well with the behavior observed in our model. These observations strengthen our belief that the phenomena and the sensitivities described in this paper are real.

Acknowledgments. Many thanks are due to the members of the KNMI predictability group for creating a stimulating working environment. Among them a few should be mentioned in particular: T. Opsteegh and R. Pasmanter for help and advice during this research and N. Weber for carefully reading this document. Finally, the KNMI studio is acknowledged for editing the figures. This work has partly been supported by the working group on Meteorology and Physical Oceanography (MFO) of the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Bryan, F., 1986a: Maintenance and variability of the thermohaline circulation. Ph.D. thesis, Princeton University, 254 pp.

——, 1986b: High latitude salinity effects and interhemisheric thermohaline circulation. *Nature,* **323,** 301–304.

Bryan, K., and M. D. Cox, 1972: An approximate equation of state for numerical models of ocean circulation. *J. Phys. Oceanogr.*, **2,** 510–514.

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

Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadel variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Climate,* **6,** 1993–204.

Duplessy, J. C., N. J. Shackleton, R. G. Fairbanks, L. Labeyrie, D. Oppo, and N. Kallel, 1988: Deepwater source variations during the last climatic cycle and their impact on the global deepwater circulation. *Paleoceanogr.*, **3,** 343–360.

Hasselman, K., 1982: An ocean model for climate variability studies. *Progress in Oceanography,* Vol. II, Pergamon, 69–92.

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

Killworth, P. D., 1983: Deep convection in the deep ocean. *Rev. Geophys. Space Phys.*, **21,** 1–26.

——, 1985: A two level wind and buoyancy driven thermohaline model. *J. Phys. Oceanogr.*, **15,** 1414–1432.

Lehman, S. J., and L. D. Keigwin, 1992: Sudden changes in the North Atlantic circulation during the last deglaciation. *Nature,* **356,** 757–762.

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

Marotzke, J., 1990: Instabilities and multiple equilibria for the thermohaline circulation. Ph.D. thesis, Institut für Meereskunde, Kiel, 126 pp.

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

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

Power, S. B., 1994: Climate drift in a global ocean circulation model. *J. Phys. Oceanogr.*, **24,** submitted.

——, and R. Kleeman, 1993: Multiple equilibria in a global ocean general circulation model. *J. Phys. Oceanogr.*, **23,** 1670–1681.

——, and ——, 1994: Surface heat flux parameterization and the response of OGCMs to high latitude freshening. *Tellus,* in press.

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

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

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

UNESCO, 1981: Tenth report of the joint panel on oceanographic tables and standards. UNESCO Tech. Papers in Marine Science, Vol. 36, UNESCO, Paris, 25 pp.

Veum, T., E. Jansen, M. Arnold, I. Beyer, and J. C. Duplessy, 1992: Water mass exchange between the North Atlantic and the Norwegian Sea during the past 28,000 years. *Nature,* **356,** 783–785.

Weaver, A. J., and E. S. Sarachik, 1991: Evidence for decadal variability in an ocean general circulation model: An advective mechanism. *Atmos.-Ocean,* **29,** 197–231.

——, J. Marotzke, P. F. Cummins, and E. S. Sarachik, 1993: Stability and variability of the thermohaline circulation. *J. Phys. Oceanogr.*, **1,** 39–60.

Welander, P., 1982: A simple heat salt oscillator. *Dyn. Atmos. Oceans,* **6,** 233–242.

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

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