A bifurcation study of the three-dimensional thermohaline ocean circulation: The double hemispheric case

by Wilbert Weijer[1] and Henk A. Dijkstra[1]

ABSTRACT

Within a low-resolution primitive-equation model of the three-dimensional ocean circulation, a bifurcation analysis is performed of double-hemispheric basin flows. Main focus is on the connection between results for steady two-dimensional flows in a nonrotating basin and those for three-dimensional flows in a rotating basin. With the use of continuation methods, branches of steady states are followed in parameter space and their linear stability is monitored. There is a close qualitative similarity between the bifurcation structure of steady-state solutions of the two- and three dimensional flows. In both cases, symmetry-breaking pitchfork bifurcations are central in generating a multiple equilibria structure. The locations of these pitchfork bifurcations in parameter space can be characterized through a zero of the tendency of a particular energy functional. Although balances controlling the steady-state flows are quantitatively very different, the zonally averaged patterns of the perturbations associated with symmetry-breaking are remarkably similar for two-dimensional and three-dimensional flows, and the energetics of the symmetry-breaking mechanism is in essence the same.

1. Introduction

Transitions between different ocean circulation patterns are one of the potential sources of (past) climate changes (Broecker *et al.,* 1985). As of now, there are many indications that the ocean circulation in the North Atlantic was different in the past. For example, during the Last Glacial Maximum the zonally-averaged northward flow likely was weaker than today, with a correspondingly smaller poleward heat transport (Bradley, 1999). It is, therefore, important to understand the equilibria of the ocean circulation under given forcing conditions and to determine their sensitivity to perturbations. This knowledge will also provide clues on the likelihood of the occurrence of rapid transitions from the present state (Tziperman, 2000).

For some time, it has been known that an ocean model forced by two agents controlling the surface buoyancy flux (heat and freshwater) allows for different circulation patterns under the same forcing conditions (Stommel, 1961; Welander, 1986). In a representation of a single-hemispheric ocean flow, for example in a two-box model, forcing conditions can

---

1. Institute for Marine and Atmospheric Research, Department of Physics and Astronomy, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. *email: w.weijer@phys.uu.nl*

be found for which a flow with sinking at the equator and one with sinking at the pole exist simultaneously. Different restoring times for salt and heat at the ocean-atmosphere interface—as imposed through the boundary conditions—are central to the occurrence of these equilibria. The limit of prescribed temperature and freshwater flux is referred to as mixed boundary conditions, whereas the limit of both prescribed salinity and temperature is referred to as restoring conditions.

The issue of multiple equilibria was addressed in a double-hemispheric set-up of the Princeton General Circulation Model (GCM) by Bryan (1986). The solution of a single-hemispheric version of the model (over the domain 0–90N), obtained under restoring conditions with observed salinity, is reflected through the equator to get a symmetric state with sinking at both poles. The freshwater flux is diagnosed from this solution, and mixed boundary conditions are applied over the whole domain. When a negative salinity anomaly of 1 psu is (instantaneously) added poleward of 45S, an asymmetric state is reached within 50 years with sinking in the north. Adding a positive salinity anomaly of 2 psu in the same region (poleward of 45S) leads to an asymmetric southern sinking state in about 200 years.

Over the last decade, many studies have tried to explain the results of Bryan (1986) within a simpler model set-up. For a double-hemispheric basin, multiple flow patterns occur under equatorially symmetric mixed boundary conditions within a large class of models. In box models (Welander, 1986) and two-dimensional Boussinesq models (Thual and McWilliams, 1992; Quon and Ghil, 1992; Cessi and Young, 1992; Dijkstra and Molemaker, 1997) these multiple equilibria arise through symmetry-breaking pitchfork bifurcations. In addition to symmetric solutions with sinking at both poles (called TH states) or solely at the equator (called SA states), also pole to pole solutions exist, the latter having either southern (SPP states) or northern (NPP states) sinking. Note that in these models, the effect of rotation is completely ignored and dominant momentum balances in the flow are between buoyancy forcing and friction. In zonally-averaged models, where the effect of rotation is somehow parameterized, transitions between TH and NPP/SPP equilibria were found as well (Wright and Stocker, 1991). Also in these cases, a pitchfork bifurcation is at the origin of the multiple equilibria (Vellinga, 1996) and the bifurcation diagram is qualitatively similar to that for the two-dimensional Boussinesq models.

In Klinger and Marotzke (1999), an attempt is made to determine bifurcation diagrams of the three-dimensional double-hemispheric configuration by calculating steady states within a low-resolution GCM for many values of parameters. A clever way is found to determine asymmetric states under equatorially symmetric conditions by independently varying the temperature differences over the northern and southern part of the flow domain. Several equilibria are found in the symmetric double-hemispheric configuration. For the case when the equator-to-pole temperature difference is the same in both hemispheres, the structure of equilibria appears to arise through a (subcritical) pitchfork bifurcation. A case with a smaller temperature difference in the northern hemisphere leads to several different asymmetric states and the bifurcation diagram is more complicated. Although precise statements on the bifurcation structure and symmetry-breaking cannot be obtained through

a number of time-dependent simulations (since the unstable steady states will always remain hidden), the results suggest a good qualitative correspondence with those in two-dimensional models.

It is only very recently that techniques of numerical bifurcation theory have reached the stage to be able to handle three-dimensional thermohaline ocean flows (Dijkstra *et al.*, 2001). We use these techniques here to study the relation between the equilibrium solutions in two-dimensional and three-dimensional models. Starting at essentially two-dimensional results as in Dijkstra and Molemaker (1997), the effect of rotation, wind forcing and convective adjustment on the bifurcation diagram is studied systematically. Main contributions of the paper are (i) to demonstrate the qualitative similarity of the structure of steady solutions in the two- and three-dimensional cases under mixed boundary conditions, and (ii) to characterize the locations of the pitchfork bifurcation responsible for the multiple equilibria energetically. Using this characterization we show that the physical mechanism of symmetry-breaking is essentially the same in two- and three-dimensional cases, although the balances controlling the steady flow are substantially different.

2. The ocean model

a. Equations and forcing

Consider a sector flow domain $[\lambda_W, \lambda_E] \times [\phi_S, \phi_N]$ representing an ocean basin on a sphere with radius r_0. The basin has constant depth D, and is rotating with angular velocity $\Omega = \eta_f \Omega_0$, where Ω_0 is the rotation rate of the earth and $\eta_f \in [0, 1]$ is a dimensionless parameter controlling the rate of rotation. The ocean velocities in eastward and northward directions are indicated by u and v, the vertical velocity is indicated by w, the pressure by p and the temperature and salinity by T and S, respectively. Vertical and horizontal mixing of momentum and of heat and salt is represented by eddy diffusivities, with horizontal and vertical friction coefficients A_H and A_V for momentum and horizontal and vertical diffusivities K_H and K_V for heat. The mixing coefficients of salt are taken equal to those of heat. A linear equation of state is taken with expansion coefficients α_T and α_S, reference temperature T_0, salinity S_0 and density ρ_0. The governing equations, using the shallow layer approximation $D/r_0 \ll 1$, are

$$\frac{Du}{dt} - \frac{uv \tan \phi}{r_0} - 2\eta_f \Omega_0 v \sin \phi = -\frac{1}{\rho_0 r_0 \cos \phi} \frac{\partial p}{\partial \lambda} + A_V \frac{\partial^2 u}{\partial z^2}$$
$$+ A_H \left(\nabla_H^2 u - \frac{u}{r_0^2 \cos^2 \phi} - \frac{2 \sin \phi}{r_0^2 \cos^2 \phi} \frac{\partial v}{\partial \lambda} \right) \tag{1a}$$

$$\frac{Dv}{dt} + \frac{u^2 \tan \phi}{r_0} + 2\eta_f \Omega_0 u \sin \phi = -\frac{1}{\rho_0 r_0} \frac{\partial p}{\partial \phi} + A_V \frac{\partial^2 v}{\partial z^2}$$
$$+ A_H \left(\nabla_H^2 v - \frac{v}{r_0^2 \cos^2 \phi} + \frac{2 \sin \phi}{r_0^2 \cos^2 \phi} \frac{\partial u}{\partial \lambda} \right) \tag{1b}$$

$$\frac{\partial p}{\partial z} = -\rho g \tag{1c}$$

$$0 = \frac{\partial w}{\partial z} + \frac{1}{r_0 \cos \phi} \left(\frac{\partial u}{\partial \lambda} + \frac{\partial (v \cos \phi)}{\partial \phi} \right) \tag{1d}$$

$$\frac{DT}{dt} = \nabla_H \cdot (K_H \nabla_H T) + \frac{\partial}{\partial z} \left(K_V \frac{\partial T}{\partial z} \right) \tag{1e}$$

$$\frac{DS}{dt} = \nabla_H \cdot (K_H \nabla_H S) + \frac{\partial}{\partial z} \left(K_V \frac{\partial S}{\partial z} \right) \tag{1f}$$

$$\rho = \rho_0 (1 - \alpha_T (T - T_0) + \alpha_S (S - S_0)) \tag{1g}$$

with:

$$\frac{DF}{dt} = \frac{\partial F}{\partial t} + \frac{u}{r_0 \cos \phi} \frac{\partial F}{\partial \lambda} + \frac{v}{r_0} \frac{\partial F}{\partial \phi} + w \frac{\partial F}{\partial z}$$

$$\nabla_H F = \left(\frac{1}{r_0 \cos \phi} \frac{\partial F}{\partial \lambda}, \frac{1}{r_0} \frac{\partial F}{\partial \phi} \right)$$

$$\nabla_H \cdot \mathbf{F} = \frac{1}{r_0 \cos \phi} \frac{\partial}{\partial \lambda} F^\lambda + \frac{1}{r_0 \cos \phi} \frac{\partial}{\partial \phi} (F^\phi \cos \phi)$$

where F and $\mathbf{F} = (F^\lambda, F^\phi)$ are an arbitrary scalar and vector, respectively.

The ocean circulation is driven by a wind stress $\boldsymbol{\tau}(\lambda, \phi) = \tau_0(\tau^\lambda, \tau^\phi)$, where τ_0 is the amplitude and $\tau^\lambda(\lambda, \phi)$ and $\tau^\phi(\lambda, \phi)$ provide the spatial pattern. The thermohaline component of the circulation is driven by heat and freshwater fluxes at the surface. The downward heat flux Q_{oa} is assumed proportional to the temperature difference between the ocean surface temperature and a prescribed atmospheric temperature $T_S(\lambda, \phi)$; i.e., $Q_{oa} = B_T(\eta_T T_S - T)$, with B_T being an interfacial exchange coefficient of heat (Haney, 1971) and η_T a dimensionless parameter controlling the amplitude of T_S. The freshwater flux is converted to an equivalent salt flux and is simply a prescribed (dimensionless) function $F_S(\lambda, \phi)$ with dimensional amplitude F_0.

The transfer of heat, freshwater and momentum from the surface downward occurs in thin boundary layers; for example in an Ekman layer, which cannot be resolved explicitly. Hence, as in low-resolution GCMs, the surface forcing is distributed as a body forcing over a certain depth of the upper ocean, using a vertical profile function $g(z)$. More explicitly, the right-hand side of the horizontal momentum (1a, 1b), temperature (1e) and salinity (1f) equations are extended with the source terms

$$Q_\tau^\lambda = g(z) \frac{\tau_0}{\rho_0 H_m} \tau^\lambda; \quad Q_\tau^\phi = g(z) \frac{\tau_0}{\rho_0 H_m} \tau^\phi \tag{2a}$$

$$Q_T = g(z) \, \frac{\eta_T T_S - T}{\tau_T}; \; Q_S = g(z) \, \frac{F_0}{H_m} F_S \tag{2b}$$

where H_m is a typical vertical scale of variation of the function $g(z)$ and τ_T is a restoring time scale to the atmospheric forcing. Using these source terms, the appropriate boundary conditions for temperature, salinity and wind stress at the ocean-atmosphere boundary are no-flux conditions. This guarantees, for example, that the surface integral of the heat flux is zero for each steady solution (Weaver and Hughes, 1996).

A nondimensional temperature $\hat{T}$, salinity $\hat{S}$ and pressure $\hat{p}$ are introduced through $T = T_0 + \Delta T \hat{T}$, $S = S_0 + \Delta S \hat{S}$ and $p = -\rho_0 gz + 2\Omega_0 r_0 U \rho_0 \hat{p}$. A characteristic horizontal velocity is indicated by U, and the governing equations are further nondimensionalized using scales r_0, D, U, DU/r_0 and r_0/U for horizontal length, vertical length, horizontal velocity, vertical velocity and time, respectively. They become

$$\epsilon_R \left(\frac{Du}{dt} - uv \tan \phi \right) - \eta_f v \sin \phi = -\frac{1}{\cos \phi} \frac{\partial p}{\partial \lambda} + E_V \frac{\partial^2 u}{\partial z^2} + \alpha_\tau \tau^\lambda g(z)$$
$$+ E_H \left(\nabla_H^2 u - \frac{u}{\cos^2 \phi} - \frac{2 \sin \phi}{\cos^2 \phi} \frac{\partial v}{\partial \lambda} \right) \tag{3a}$$

$$\epsilon_R \left(\frac{Dv}{dt} + u^2 \tan \phi \right) + \eta_f u \sin \phi = -\frac{\partial p}{\partial \phi} + E_V \frac{\partial^2 v}{\partial z^2} + \alpha_\tau \tau^\phi g(z)$$
$$+ E_H \left(\nabla_H^2 v - \frac{v}{\cos^2 \phi} + \frac{2 \sin \phi}{\cos^2 \phi} \frac{\partial u}{\partial \lambda} \right) \tag{3b}$$

$$\frac{\partial p}{\partial z} = Ra(T - \lambda S) \tag{3c}$$

$$0 = \frac{\partial w}{\partial z} + \frac{1}{\cos \phi} \left(\frac{\partial u}{\partial \lambda} + \frac{\partial (v \cos \phi)}{\partial \phi} \right) \tag{3d}$$

$$\frac{DT}{dt} = \nabla_H \cdot (P_H \nabla_H T) + \frac{\partial}{\partial z} \left(P_V \frac{\partial T}{\partial z} \right) + B(\eta_T T_S - T) g(z) \tag{3e}$$

$$\frac{DS}{dt} = \nabla_H \cdot (P_H \nabla_H S) + \frac{\partial}{\partial z} \left(P_V \frac{\partial S}{\partial z} \right) + \gamma F_S g(z) \tag{3f}$$

where the hats are dropped for convenience. On the lateral walls, slip conditions are prescribed to allow for two-dimensional solutions in particular cases, and the heat and salt fluxes are zero. The bottom of the ocean $z = -1$ is assumed to be flat, isolated and impermeable to salt. The nondimensional boundary conditions are hence formulated as

$$z = -1, 0 : \frac{\partial u}{\partial z} = \frac{\partial v}{\partial z} = w = \frac{\partial T}{\partial z} = \frac{\partial S}{\partial z} = 0 \tag{4a}$$

$$\lambda = \lambda_W, \lambda_E : u = \frac{\partial v}{\partial \lambda} = \frac{\partial w}{\partial \lambda} = \frac{\partial T}{\partial \lambda} = \frac{\partial S}{\partial \lambda} = 0 \tag{4b}$$

$$\phi = \phi_S, \phi_N : \frac{\partial u}{\partial \phi} = v = \frac{\partial w}{\partial \phi} = \frac{\partial T}{\partial \phi} = \frac{\partial S}{\partial \phi} = 0. \tag{4c}$$

Note that the model formulated here does not guarantee stably stratified solutions. As in all other large-scale ocean models the effect of convection, which occurs when the stratification is not statically stable must be explicitly parameterized. Such a parameterization is usually referred to as 'convective adjustment' and its effects on the results is explicitly considered in Section 3d.

The parameters in the set of equations (3) are the Rossby number ϵ_R, the Rayleigh number Ra, the vertical and horizontal Ekman numbers E_V and E_H, the wind stress coefficient α_τ, the buoyancy ration λ, the vertical and horizontal inverse Péclet numbers P_V and P_H, the Biot number B and the freshwater flux strength γ. Expressions for these parameters are

$$\epsilon_R = \frac{U}{2\Omega_0 r_0}; E_V = \frac{A_V}{2\Omega_0 D^2}; E_H = \frac{A_H}{2\Omega_0 r_0^2}; \alpha_\tau = \frac{\tau_0}{2\Omega_0 \rho_0 H_m U}; \lambda = \frac{\alpha_S \Delta S}{\alpha_T \Delta T};$$

$$P_V = \frac{K_V r_0}{U D^2}; P_H = \frac{K_H}{U r_0}; B = \frac{r_0}{U \tau_T}; \gamma = \frac{F_0 r_0}{U \Delta S H_m}. \tag{5}$$

When rotation is present ($\eta_f = 1$), a geostrophic balance between the pressure gradient and the Coriolis force is expected, and the Rayleigh number Ra can be cast in its geostrophic form, Ra_g:

$$Ra = Ra_g = \frac{\alpha_T \Delta T g D}{2\Omega_0 U r_0}. \tag{6}$$

In absence of rotation, however, pressure forces are balanced by (horizontal) friction, and the Rayleigh number effectively becomes $Ra_f = Ra/E_H = Ra_g/E_H$. This means that the velocities increase characteristically by a factor $1/E_H$ when the rotation rate is reduced to zero. In order to facilitate comparison between the two-dimensional and three-dimensional model, the Rayleigh number is expressed as:

$$Ra = [\eta_f + (1 - \eta_f)E_H]Ra_g. \tag{7}$$

For $\eta_f = 0$ the Rayleigh number is rescaled by a factor E_H to compensate for the frictional balance, whereas for $\eta_f = 1$ Ra attains its geostrophic value (6).

Apart from the parameters η_T and η_f, the system appears to contain 10 parameters. However, only 8 of these are independent; when the salt field is rescaled by a factor λ, the product $\sigma = \lambda\gamma$ appears, which is an independent parameter. Moreover, the velocity scale U can be chosen as a function of other parameters, reducing the number of parameters again by one.

The continuous steady equations are invariant with respect to reflection around the equator, which is represented by

$$u(\lambda, -\phi, z) = u(\lambda, \phi, z); \ v(\lambda, -\phi, z) = -v(\lambda, \phi, z); \ w(\lambda, -\phi, z) = w(\lambda, \phi, z)$$

$$p(\lambda, -\phi, z) = p(\lambda, \phi, z); \ T(\lambda, -\phi, z) = T(\lambda, \phi, z); \ S(\lambda, -\phi, z) = S(\lambda, \phi, z). \tag{8}$$

The equations are discretized in space using a control volume discretization method on a staggered (Marker and Cell or Arakawa C−) grid, that places the p, T and S points in the center of a grid cell, and the u, v, and w points on its boundaries. This method is second-order accurate, and preserves the reflective symmetry of the system. The spatially discretized model equations can be written in the form

$$\mathbf{M} \frac{d\mathbf{u}}{dt} = \mathbf{F}(\mathbf{u}) = \mathbf{L}(\mathbf{u}) + \mathbf{N}(\mathbf{u}, \mathbf{u}) \tag{9}$$

where the vector $\mathbf{u}$ contains the unknowns (u, v, w, p, T, S) at each grid point. The operators $\mathbf{M}$ and $\mathbf{L}$ are linear and $\mathbf{N}$ represents the nonlinear terms in the equations. Steady solutions of this system are calculated, and branches of steady states are followed through parameter space using the technique of pseudo-arclength continuation. The stability of the solutions is considered by performing a linear stability analysis; the Jacobi-Davidson QZ-method is used to determine the eigenvectors and the associated eigenvalues $\varsigma = \varsigma_r + i\varsigma_i$ (where ς_r and ς_i are the real and imaginary parts, respectively). Transitions that mark qualitative changes, such as transitions to multiple equilibria (pitchfork or saddle node bifurcations) or periodic behavior (Hopf bifurcations), can be detected. More details on the continuation method and the linear stability analysis are given in Appendix A.

b. The double-hemispheric basin set-up

The domain chosen is comparable in size to the Atlantic: it is a sector of 64° wide (with $\lambda_W = 286°$ and $\lambda_E = 350°$) and 120° long (with $\phi_S = -60°$ and $\phi_N = 60°$), and has a constant depth $D = 4000$ m. A resolution of $d\lambda \times d\phi \times dz = 4° \times 3.75° \times 250$ m is chosen, which is accomplished by a grid of $N \times M \times L = 16 \times 32 \times 16$ boxes. For a similar parameter regime, a comparable resolution in a single-hemispheric basin turned out to be sufficient to reproduce results obtained with higher resolution with reasonable accuracy (Dijkstra *et al.,* 2001).

The wind stress forcing considered is an idealized profile for the Atlantic mimicking a double-gyre wind stress in each hemisphere, i.e. in dimensionless form

$$\tau^\lambda(\phi) = -\cos 4\pi \frac{\phi - \phi_S}{\phi_N - \phi_S} \tag{10a}$$
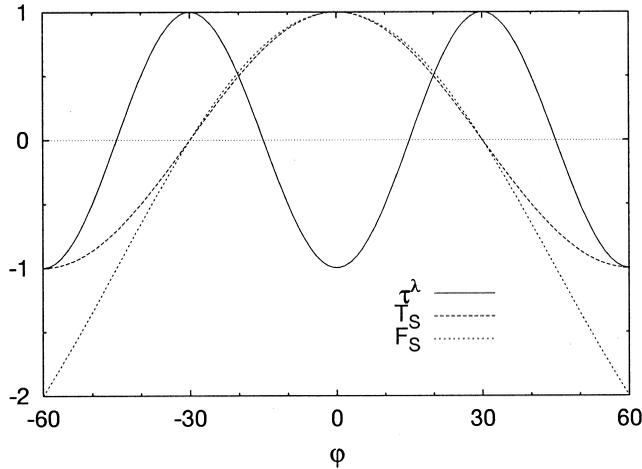
$$\tau^\phi = 0. \tag{10b}$$

Figure 1. The forcing functions for wind stress τ^λ, the atmospheric temperature T_S and the freshwater flux F_S.

The surface temperature T_S and surface freshwater flux F_S are prescribed, similar to Klinger and Marotzke (1999), as

$$T_S(\phi) = \cos \pi \frac{\phi}{\phi_N} \tag{11a}$$

$$F_S(\phi) = \frac{1}{\cos \phi} \cos \pi \frac{\phi}{\phi_N}. \tag{11b}$$

The temperature profile, the zonal wind stress and the freshwater flux are shown in Figure 1 and need no further discussion. Note that, because of the introduction of the parameter η_T, the dimensional equator-pole temperature difference over the sector is equal to $2\eta_T \Delta T$. The freshwater forcing is such that the integral over the surface is zero, which is a necessary condition for the existence of steady-state solutions.

The function $g(z)$, appearing in Eqs. (3), is chosen as

$$g(z) = \mathcal{H}(z - z_{L-1}; \epsilon_H) \tag{12}$$

where z_{L-1} is the bottom level of the uppermost grid cell, and $\mathcal{H}$ is a continuous approximation to the Heaviside function. For the latter we use

$$\mathcal{H}(x; \epsilon_H) = \frac{1}{2} \left(1 + \tanh \frac{x}{\epsilon_H} \right) \tag{13}$$

where $\epsilon_H = 10^{-6}$. In this way, the input of each quantity through the ocean-atmosphere surface (zonal and meridional momentum, heat and salt) is distributed as a source term over the uppermost level.

Table 1. Standard values of parameters used in the numerical calculations.

$2\Omega_0 = 1.4^6 \cdot 10^{-4}$	$[\text{s}^{-1}]$	$r_0 = 6.4 \cdot 10^6$	$[\text{m}]$
$\tau_0 = 1.0 \cdot 10^{-1}$	$[\text{N m}^{-2}]$	$D = 4.0 \cdot 10^3$	$[\text{m}]$
$F_0 = 1.0 \cdot 10^{-7}$	$[\text{psu m s}^{-1}]$	$U = 1.0 \cdot 10^{-1}$	$[\text{m s}^{-1}]$
$\rho_0 = 1.0 \cdot 10^3$	$[\text{kg m}^{-3}]$	$\Delta T = 1.0$	$[\text{K}]$
$\alpha_T = 1.0 \cdot 10^{-4}$	$[\text{K}^{-1}]$	$\tau_T = 75$	$[\text{days}]$
$A_H = 16.0 \cdot 10^6$	$[\text{m}^2\text{s}^{-1}]$	$A_V = 1.0 \cdot 10^{-3}$	$[\text{m}^2\text{s}^{-1}]$
$K_H = 1.0 \cdot 10^3$	$[\text{m}^2\text{s}^{-1}]$	$K_V = 1.0 \cdot 10^{-4}$	$[\text{m}^2\text{s}^{-1}]$
$\Delta S = 1.0$	$[\text{psu}]$	$\alpha_S = 7.6 \cdot 10^{-4}$	$[\text{psu}^{-1}]$
		$S_0 = 35.0$	$[\text{psu}]$
$g = 9.8$	$[\text{m s}^{-2}]$	$T_0 = 15.0$	$[\text{K}]$
$H_m = 250$	$[\text{m}]$		
$Ra = 4.2 \cdot 10^{-2}$		$E_H = 2.7 \cdot 10^{-3}$	
$\lambda = 7.6$		$E_V = 4.3 \cdot 10^{-7}$	
$\gamma = 2.6 \cdot 10^{-2}$		$P_H = 1.5 \cdot 10^{-3}$	
$\eta_T = 10.0$		$P_V^0 = 3.9 \cdot 10^{-4}$	
$B = 10.0$		$P_V^c = 1.0 \cdot 10^9$	
$\epsilon_R = 0$		$\alpha_\tau = 2.9 \cdot 10^{-2}$	

Most of the parameters are fixed at values as used in low-resolution GCMs, and these values are listed in Table 1. For these parameter values, the Rossby number ϵ_R is small $(\mathbb{O}(10^{-4}))$ and is set to zero in our calculations. Note that in the standard case, the horizontal friction coefficient A_H is rather large. The value of A_H is bounded from below by the boundary layers which develop near the continents. Near the western boundary, the Munk frictional boundary layer thickness at latitude ϕ_0 scales with $(A_H/\beta_0)^{1/3}$, where $\beta_0 = 2\Omega_0 \cos \phi_0/r_0$ monitors the variation of the Coriolis parameter. With a typical horizontal resolution of $4°$, this leads to a typical lower bound of $A_H = 5.0 \times 10^5 \text{ m}^2\text{s}^{-1}$ at $\phi_0 = 45°$. However, the thickness of the Ekman layers near the continental walls has a typical width of $(A_H/f_0)^{1/2}$, where $f_0 = 2\Omega_0 \sin \phi_0$, which restricts the value of A_H to be larger than $1.0 \times 10^7 \text{ m}^2\text{s}^{-1}$. In typical ocean models, values much smaller are taken, but it has been shown that this leads to numerical waves near these boundaries (Winton, 1996; Killworth, 1985) which show up as wiggles in the steady-state solutions (Dijkstra *et al.,* 2001).

3. Results

In the results below, steady states are computed as a function of the strength of the freshwater flux σ. By plotting a norm of the solution versus this control parameter, for every steady state computed, a so-called bifurcation diagram is obtained. As a norm, the maximum of the meridional overturning streamfunction (Ψ_M) is chosen, which is computed as follows. The nondimensional overturning streamfunction Ψ is defined by

$$\bar{v} = \frac{\partial \Psi}{\partial z}; \; \bar{w} = -\frac{\partial \Psi}{\partial \phi} \tag{14}$$
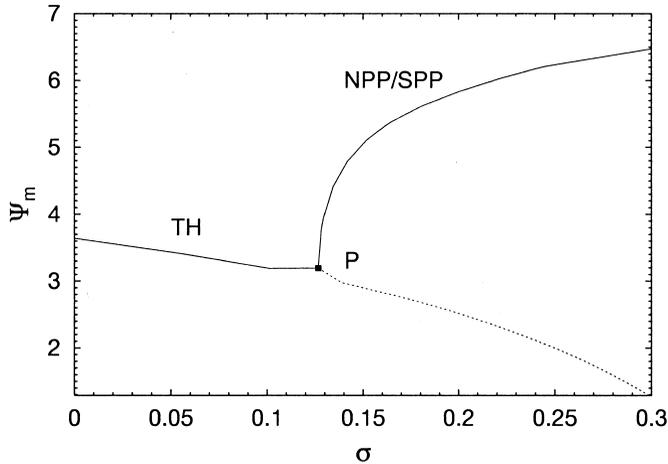
where

Figure 2. Bifurcation diagram for the two-dimensional nonrotating case $\eta_f = 0$, using $\sigma = \gamma\lambda$ as control parameter. On the vertical axis, the maximum of the meridional overturning streamfunction is plotted (in Sv). Solid lines denote branches of stable solutions, while unstable branches are dashed. For low values of σ only thermally driven solutions (TH) are possible. At the pitchfork bifurcation P this branch becomes unstable, and asymmetric pole-to-pole solutions (NPP and SPP) branch off. The latter are symmetry-related and cannot be distinguished in this plot.

$$\bar{v} = \int_{\lambda_W}^{\lambda_E} v \cos \phi \, d\lambda; \quad \bar{w} = \int_{\lambda_W}^{\lambda_E} w \cos \phi \, d\lambda. \tag{15}$$

With the scaling used, the maximum dimensional volume transport $\Psi_M = r_0 UD \max \Psi$ and this is expressed in Sv, where 1 Sv = $10^6 \, \text{m}^3\text{s}^{-1}$.

The presentation of the results in this section is as follows. First, the case in which rotation and wind-forcing are absent is considered, allowing for fully 2-dimensional solutions. Then explicitly, the transition to the rotating buoyancy-driven flows is presented and the robustness of the results with respect to the presence of wind forcing and convective adjustment is discussed.

a. Two-dimensional solutions

As conditions on the eastern and western boundary are assumed free-slip, strictly two-dimensional solutions exist in case rotation and wind forcing are absent (i.e. $\eta_f = \alpha_\tau = 0$). To connect the results here to those in earlier two-dimensional studies, we first study these two-dimensional solutions. The bifurcation diagram for $\eta_f = 0$, $\alpha_\tau = 0$ and $Ra = E_H Ra_g = 1.2 \cdot 10^{-4}$ is shown in Figure 2. For small values of σ a unique solution exists (labeled TH) that is thermally driven and has equatorial symmetry (Fig. 3a–b). This solution becomes unstable at $\sigma = 0.13$ through a super-critical pitchfork bifurcation P. At P, equatorially asymmetric circulations arise, with major downwelling either on the
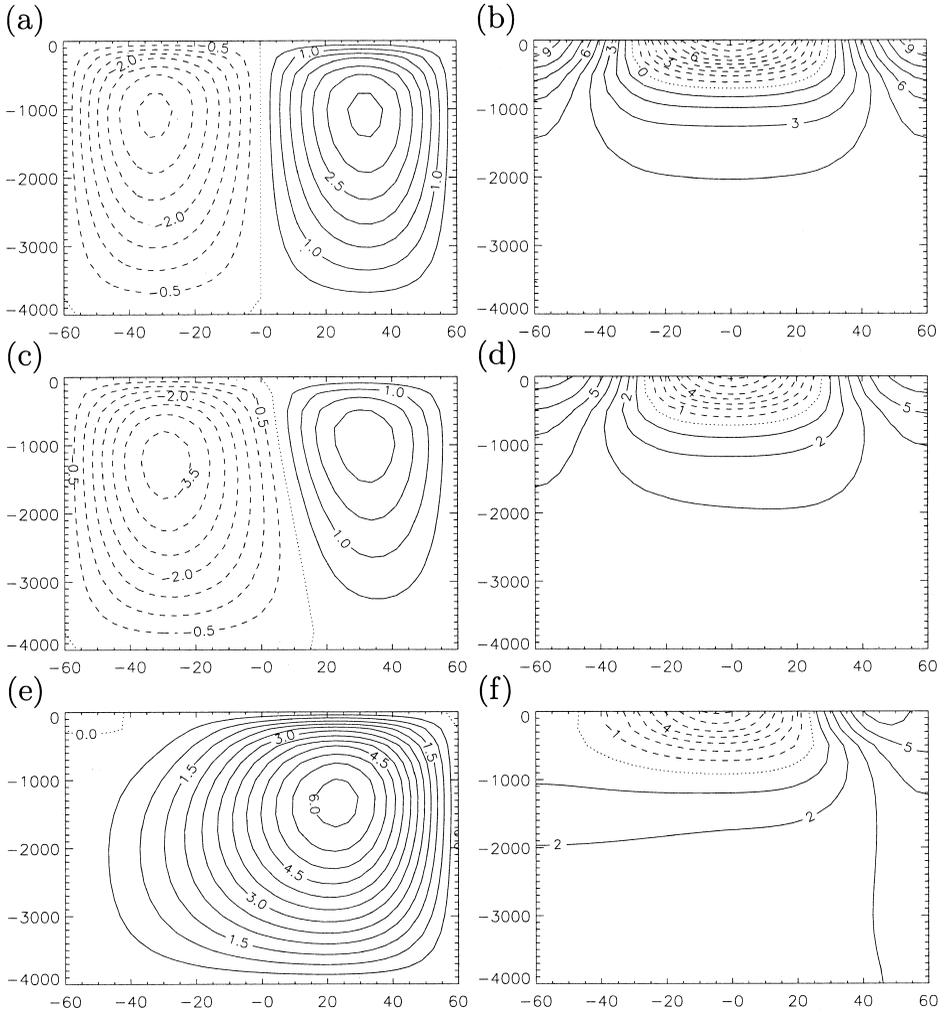
Figure 3. Overturning streamfunction Ψ and zonally averaged density ρ for the TH solution at $\sigma = 0$ (a–b), the SPP solution at $\sigma = 0.13$ (c–d), and the NPP solution at $\sigma = 0.24$ (e–f). The SPP state at $\sigma = 0.13$ is close to the pitchfork bifurcation P in Figure 2, and displays only a slight degree of asymmetry. The NPP state at $\sigma = 0.24$ shows a much stronger degree of asymmetry, and displays a clear pole-to-pole circulation. The dimensional density can be computed from $\rho^* - \rho_0 = \rho_0 \Delta T \alpha_T \rho$, as can be deduced from the scaling in Section 2.

southern hemisphere (southern sinking pole-to-pole, SPP; Fig. 3c–d) or on the northern hemisphere (northern sinking pole-to-pole, NPP; Fig. 3e–f). Note that by plotting the overturning strength Ψ_M as a norm in the bifurcation diagram, the two branches NPP and SPP cannot be distinguished: since the forcing is symmetric around the equator, both
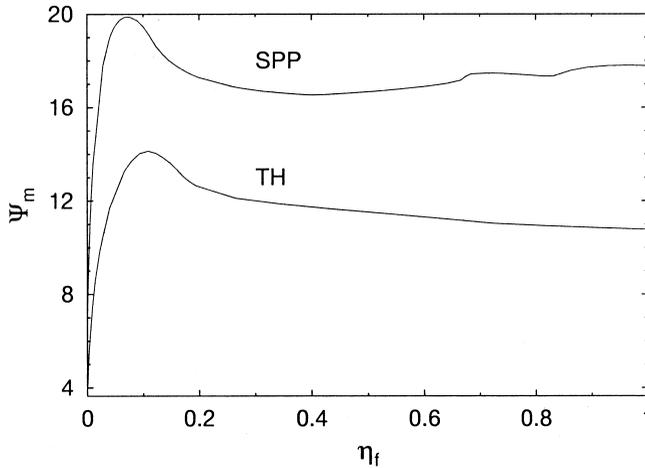
Figure 4. The response of the overturning strength upon introducing rotation for a TH solution ($\sigma = 0$) and an SPP solution ($\sigma = 0.30$). Rotation off (on) is represented by $\eta_f = 0$ ($\eta_f = 1$). Note that the Rayleigh number is rescaled according to $Ra = [\eta_f + (1 - \eta_f)E_H]Ra_g$, where Ra_g is the original 'geostrophic' form of the Rayleigh number. The curves connect the TH and SPP branches in the bifurcation diagrams of Figures 2 and 6.

solutions are symmetry related through (8) and hence their overturning strength is the same.

Clearly, the results for these two-dimensional solutions in a three-dimensional model are qualitatively very similar to the results for two-dimensional Boussinesq models in a rectangular ocean basin (Dijkstra and Molemaker, 1997). A pitchfork bifurcation P gives rise to equatorially asymmetric circulations, and is thus the origin of the multiple equilibria. Hence, the effects of the spherical geometry do not alter the qualitative aspects of the solution structure.

b. Effect of rotation

For the same values of parameters as in the previous section, the two-dimensional solutions are followed into the rotating regime by varying the dimensionless Coriolis parameter η_f from zero to unity. Simultaneously, the Rayleigh number is rescaled according to (7), and attains its original 'geostrophic' value for $\eta_f = 1$.

The response of the overturning strength of an SPP solution (at $\sigma = 0.30$) and a TH solution (at $\sigma = 0.0$) is shown in Figure 4. It is influenced by two effects: the initial increase in overturning strength of both circulation patterns reflects the increase in the Rayleigh number (and scales according to $\Psi_M \propto Ra^{1/2}$ in both cases). The subsequent decrease in overturning strength is caused by the Coriolis term taking over from friction in balancing the pressure force, once η_f is large enough. The dominant momentum balance changes from frictionally controlled to geostrophically controlled, so that the meridional

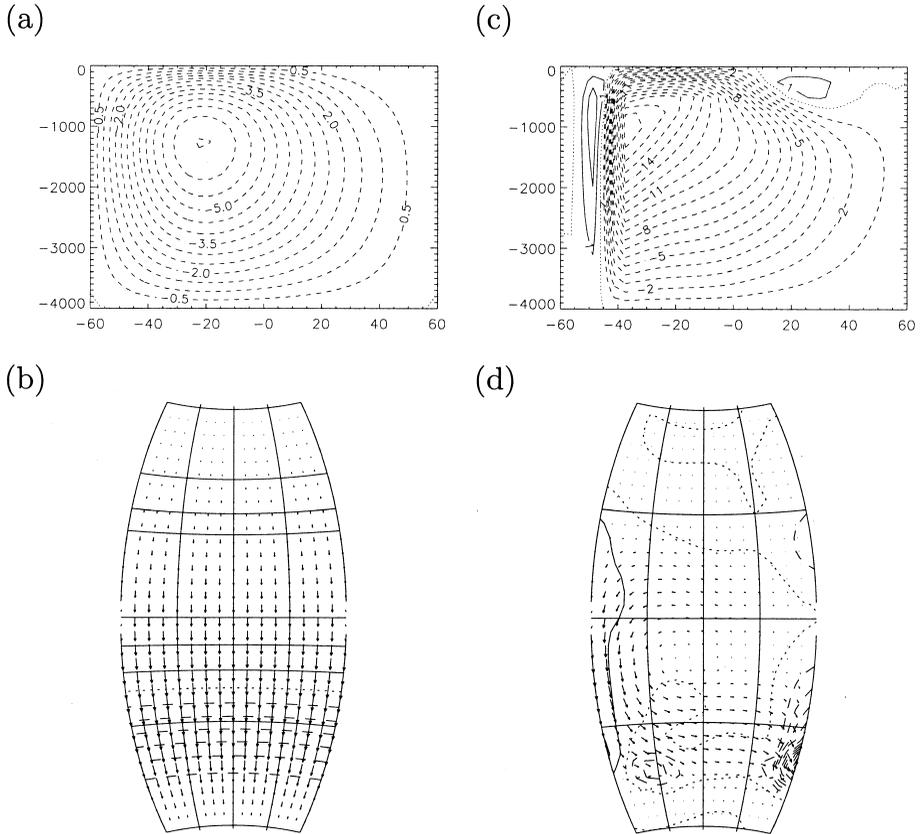(a)

(c)

(b)

(d)

Figure 5. The overturning streamfunction Ψ and the surface velocity field (vector plot of the horizontal velocity and contour plot of the vertical velocity) of the SPP circulation patterns at $\eta_f = 0$ (a–b) and $\eta_f = 1$ (c–d) in Figure 4.

pressure gradient induced by the surface forcing now generates zonal, rather than meridional, flow. These zonal transports, in turn, generate a zonal pressure gradient that ultimately drives a meridional overturning. This indirect forcing of the meridional circulation by a meridional pressure gradient is less efficient and explains the decrease in overturning strength of both the TH and SPP modes with increasing η_f.

The change of the SPP solution from the nonrotating to the rotating case is shown through the solutions at $\eta_f = 0$ and $\eta_f = 1$ in Figure 5. To represent the three-dimensional flows, the overturning streamfunctions and the velocity fields at $z = -125$ m are plotted. The zero-rotation solution (Fig. 5a–b) is the pure SPP solution at $\sigma = 0.30$. The surface velocity plot shows that it is indeed perfectly two dimensional. The overturning is weak, and the downwelling takes place in a broad region poleward of 20S. Under intermediate rotation, the flow becomes more and more zonal and the downwelling region is severely narrowed. In the full rotating case (Fig. 5c–d), a strong zonal flow has developed in the
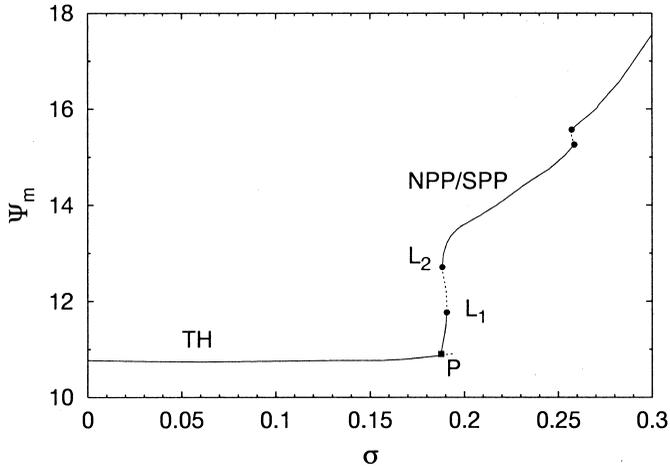
Figure 6. Bifurcation diagram in the control parameter σ for the case of full rotation ($\eta_f = 1$) and parameters similar to those of Figure 2. The PP branches lose their stability at the limit point L_1, but regain stability at L_2. The limit points close to $\sigma = 0.25$ reflect minor rearrangements of the steady state patterns.

southern hemisphere. The main downwelling has been concentrated in a small area in the southeastern corner of the model at about 45S. The bifurcation diagram for the case $\eta_f = 1$ is presented in Figure 6, again using σ as a control parameter. The overturning strength of the TH solution remains quite constant despite the increase of the freshwater flux amplitude. The solution in Fig. 5c–d is indeed on an SPP branch which arises through a super-critical pitchfork bifurcation (P) at $\sigma = 0.19$ (not far from that in the two-dimensional case). At limit point L_1, the PP branches become unstable, but regain stability through a second limit point L_2. The overturning increases strongly with σ on this branch.

Although patterns of the steady states change markedly in a rotating basin because of the strong zonal flow developing, the qualitative properties of the bifurcation diagram are robust and in agreement with that of the two-dimensional solutions. There is even a fair quantitative agreement, once the proper rescaling of the Rayleigh number is made; with the rescaling as above, the pitchfork bifurcations can be found at approximate values of σ.

c. Most unstable mode

From the linear stability analysis, the pattern of the mode destabilizing the TH state at the symmetry-breaking pitchfork bifurcation is determined from the eigenvector **x** in (A5) corresponding to the eigenvalue $\varsigma_r = \varsigma_i = 0$ at P in Figure 6. The steady state at P is plotted in Figure 7. In addition to zonally averaged profiles of T, S and ρ (panels a–c), the overturning streamfunction (panel d) and the velocity fields at 100 m and 3000 m depth (panels e and f) are shown. All fields are equatorially symmetric and the density field is stably stratified except at high latitudes. Note that due to rotation the flow is mainly zonal,
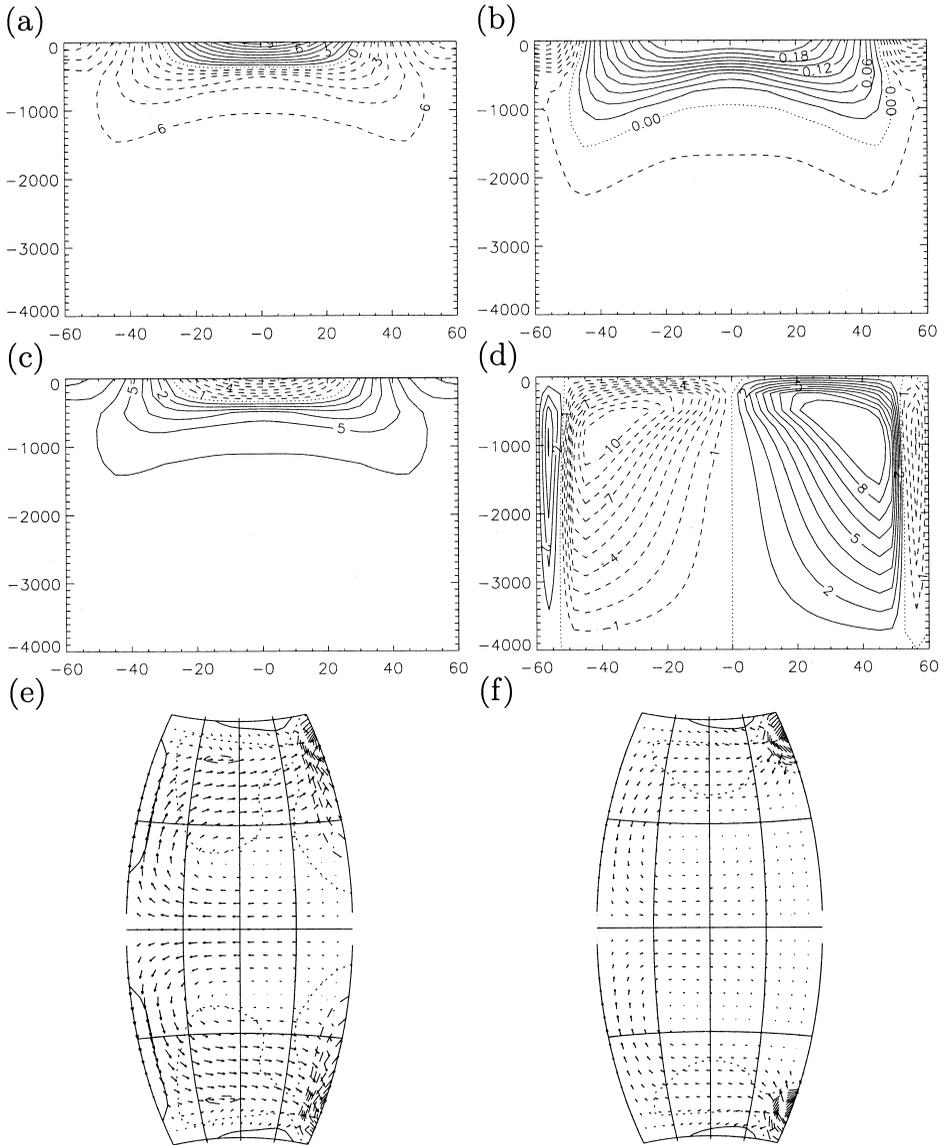
Figure 7. Steady state at the pitchfork bifurcation P in Figure 6. Shown are the zonally averaged fields of (a) temperature T, (b) salinity S, and (c) density, as well as (d) the overturning streamfunction Ψ, and the velocity fields at (e) 100 m and (f) 3000 m depth. Dimensional temperature and salinity can be computed from $T^* - T_0 = \Delta T T$ and $S^* - S_0 = \Delta S S$.

except in the western boundary layers. The main downwelling is confined to small areas in the north- and southeastern corners of the domain. The same fields for the most unstable eigenvector at the pitchfork bifurcation are shown in Figure 8. Note that, if **x** is an
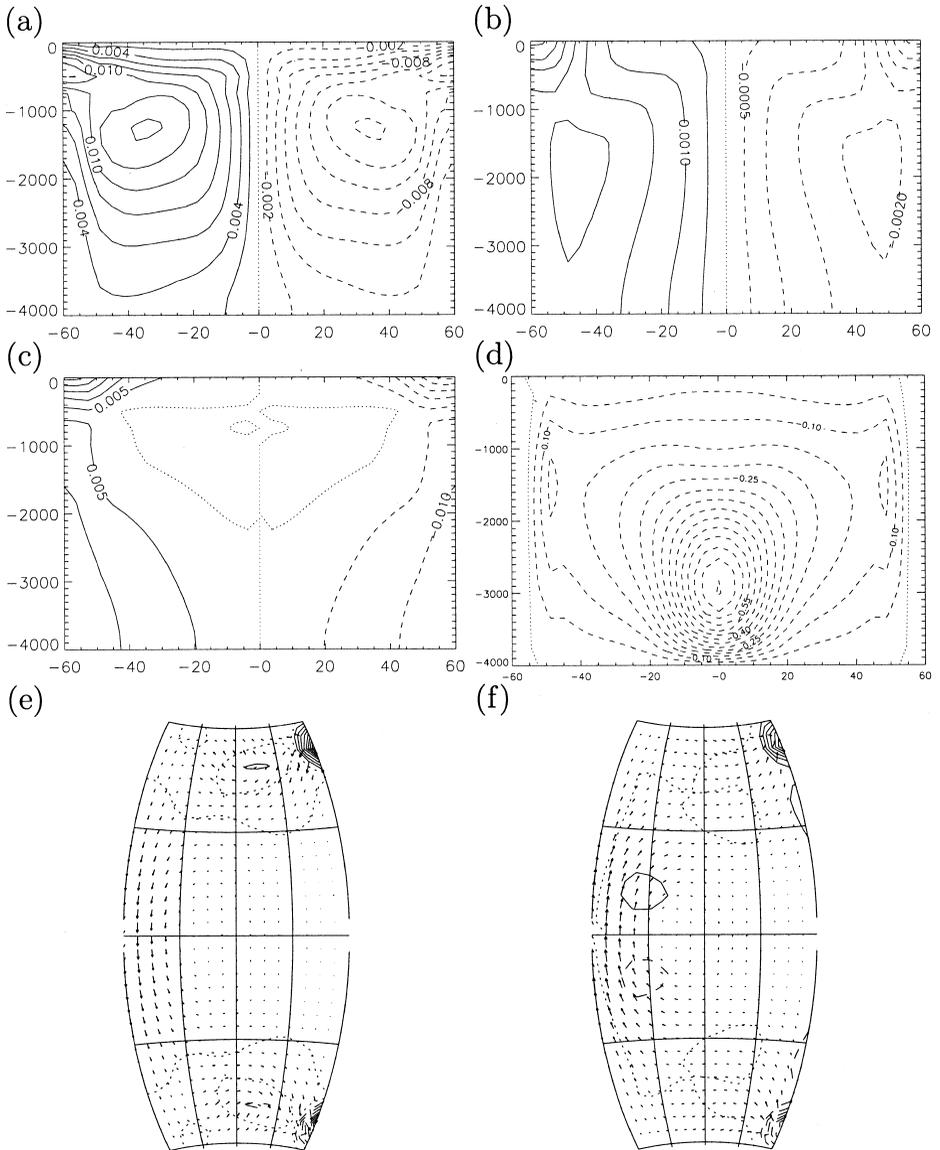
Figure 8. Most unstable eigenvector at the pitchfork bifurcation P in Figure 6, destabilizing the steady TH state (Fig. 7). Same fields as in Figure 7. Although the amplitude of the eigenvector is undetermined, the amplitudes of the individual fields are mutually consistent. The perturbation shown will favor a transition to an SPP state, but note that also the sign-reversed version of this perturbation is an eigenvector, and this will favor a transition to an NPP state.

eigenvector, $-\mathbf{x}$ is an eigenvector as well. The eigenvector plotted here will favor a transition to an SPP state.

Despite the zonal structure that results from the presence of rotation, the zonally averaged structures of the destabilizing perturbation are strikingly similar to those in purely two-dimensional models (Dijkstra and Molemaker, 1997) or zonally averaged models (Vellinga, 1996). Both the salinity and the temperature perturbations have a bi-polar structure (panels a and b), which is positive on the southern hemisphere, and negative in the north. This gives rise to an equatorially antisymmetric density perturbation (panel c). Its mainly positive sign in the southern hemisphere shows that it is dominated by salinity, although the thermal perturbation wins at mid-depth between 40S and 40N. This bi-polarity of the anomalous density field sets up an interhemispheric pressure difference at depth. This generates deep cross-equatorial flow that is confined to the viscous western boundary layer (panel f). A return flow at shallower levels is established for reasons of continuity (panel e). From this, the perturbation overturning streamfunction can be understood (panel d). Furthermore, the density at the surface and at depth increases southward and, through the thermal wind balance, this sets up a zonal flow. In the southern hemisphere, this flow is eastward at the surface and westward at depth, in the north it is the other way around (panels e and f).

The overturning perturbation advects heat and salt from (sub)tropical regions southward, and enhances the thermal and saline anomalies on that hemisphere. When the meridional salinity gradient is strong enough, the density anomaly is controlled by salinity, and the density perturbation, that initially generated the overturning anomaly, is amplified. The surface salt flux amplitude σ must, therefore, exceed a critical value for the TH circulation to become unstable.

d. Robustness: wind stress and convective adjustment

Before we explore the physics of the symmetry-breaking bifurcation in more detail, we will first have to convince ourselves that the pitchfork bifurcation is a robust feature of the three-dimensional flow. We have to show that it does not disappear when wind stress is added, or when convective adjustment is used to obtain stably stratified solutions. The bifurcation diagram for the model with full wind stress forcing (with α_τ having the standard value) is plotted in Figure 9. Comparison with Figure 6 shows that the position of the limit point L_2 has hardly changed, but that the limit point L_1 and the pitchfork bifurcation P have shifted dramatically to larger values of σ. Consequently, a large window has opened where more than two stable equilibria can be found. Between $\sigma(L_2)$ and $\sigma(P)$, three stable equilibria coexist, namely the symmetrical TH solution, and the two pole-to-pole solutions. Between $\sigma(P)$ and $\sigma(L_1)$ no less than four stable equilibria are present, the strong pole-to-pole branches of the NPP and SPP solutions, and two weaker varieties. It is clear that the TH circulation is stabilized considerably by the equatorially symmetric wind stress. This is due to the fact that the wind-induced meridional transports
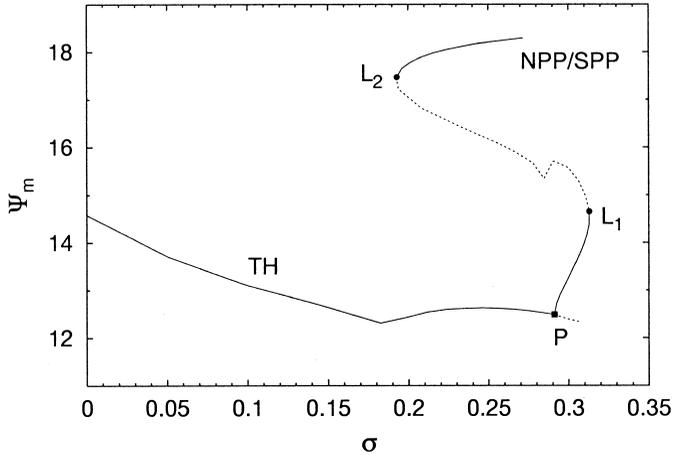
Figure 9. Bifurcation diagram for the rotational case with full wind stress. Note the shift of the pitchfork bifurcation P to larger values of σ, opening a large window where multiple equilibria exist.

modify the background density structure to such an extent that the destabilizing salt advection feedback is severely hindered.

The broadening of the region of multiple equilibria is interesting from a paleoclimatic point of view. It creates possibilities for explaining rapid transitions in the thermohaline circulation regime by switches between equilibria other than just the SPP and NPP states alone (e.g. Stocker, 2000). Admittedly, the effect of the wind-driven circulation is taken into account in a rudimentary way because of the low resolution and high friction used. But the present result shows that purely three-dimensional features that are not represented in two-dimensional models, may impact considerably on the bifurcation structure.

Since convection, which occurs in the case of an unstable stratification, is not resolved by the hydrostatic model, an explicit representation is needed to obtain stably stratified solutions. One way to do this is to increase the vertical mixing coefficients locally when the density stratification is unstable, i.e.,

$$P_V = P_V^0 + P_V^c \mathcal{H}\left(\frac{\partial \rho}{\partial z}; \epsilon_H\right) \tag{16}$$

where P_V^0 is the background inverse Péclet number, P_V^c is the convective inverse Péclet number (which is much larger than P_V^0) and $\mathcal{H}$ is a continuous approximation to the Heaviside function (13) with $\epsilon_H = 0.1$. Unfortunately, the continuation to stably stratified solutions may be very problematic. Upon increasing P_V^c for the unstably stratified solutions of, for instance, Figure 6, an endless series of limit points may be encountered that severely hinders the progression to sufficiently large values of P_V^c (cf. Vellinga, 1998). Therefore, an alternative algorithm was developed to construct a stably stratified solution $\mathbf{u}_{st}$ from a
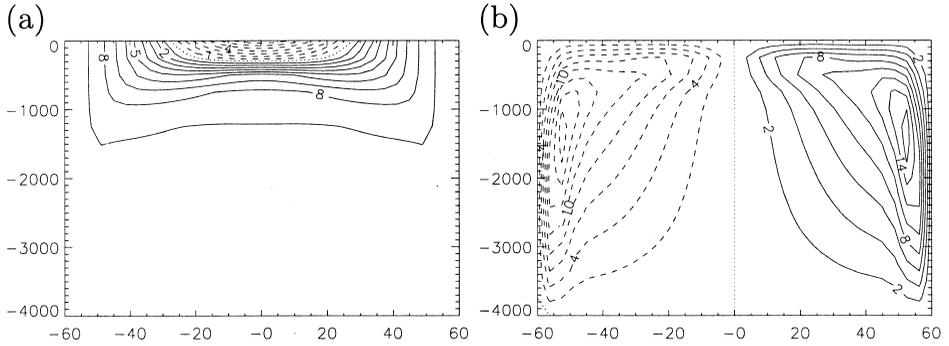
Figure 10. (a) Zonally averaged density field and (b) overturning streamfunction of a stably stratified
solution, obtained from the steady TH state at the pitchfork bifurcation P in Figure 6.

solution $\mathbf{u}$ of system (A2), the Global Adjustment Procedure (GAP). We must note that the
'irreversible' nature of the algorithm makes it impossible to continue in stably stratified
solutions, hence no bifurcation diagrams of stably stratified solutions can be constructed.
The procedure of the GAP is described in Appendix B. The Global Adjustment Procedure
was applied to the TH state at the pitchfork bifurcation P in Figure 6. In an iteration of
$N_a = 12$ steps, and with $P_V^c = 10^9$ (Table 1) we reached a stably stratified TH solution.
Figure 10a shows that the GAP has indeed removed the static instabilities of the original
solution (Fig. 7c). Moreover, the counter-rotating polar overturning cells in 7d have
disappeared (Fig. 10b). The overturning strength increased considerably, from 10.9 Sv to
17.2 Sv. Computation of the most unstable mode shows that the stably stratified solution is
unstable with respect to an asymmetric perturbation (Fig. 11) with the same bi-polar
structure as in the unstably stratified case (Fig. 8). Comparison of the spatial patterns of the
most unstable eigenvectors shows that the patterns are very much alike, although the
structure at high latitudes is clearly removed by convective mixing. This suggests that the
destabilizing mechanism of symmetry-breaking is presumably not affected by the convec-
tive adjustment procedure. In fact, for the class of stably stratified solutions the symmetry-
breaking pitchfork bifurcation is situated at lower values of σ than for the unstable
solutions; the real part of the eigenvalue $\varsigma_r > 0$ for the stably stratified eigenvector at
$\sigma = \sigma(P)$, whereas $\varsigma_r = 0$ for its unstably stratified counterpart.

4. Energetics of the instability mechanism

As can be concluded from the results above, the symmetry-breaking pitchfork bifurca-
tion is central to the connection between results of the two- and three-dimensional
solutions. Zonally averaged patterns of the destabilizing perturbations are remarkably
similar, despite the large differences in the local momentum balance. To determine a more
detailed relation between the instability mechanisms in both cases, an energy analysis is
applied. As in all instabilities, a perturbation exchanges energy with the mean state by
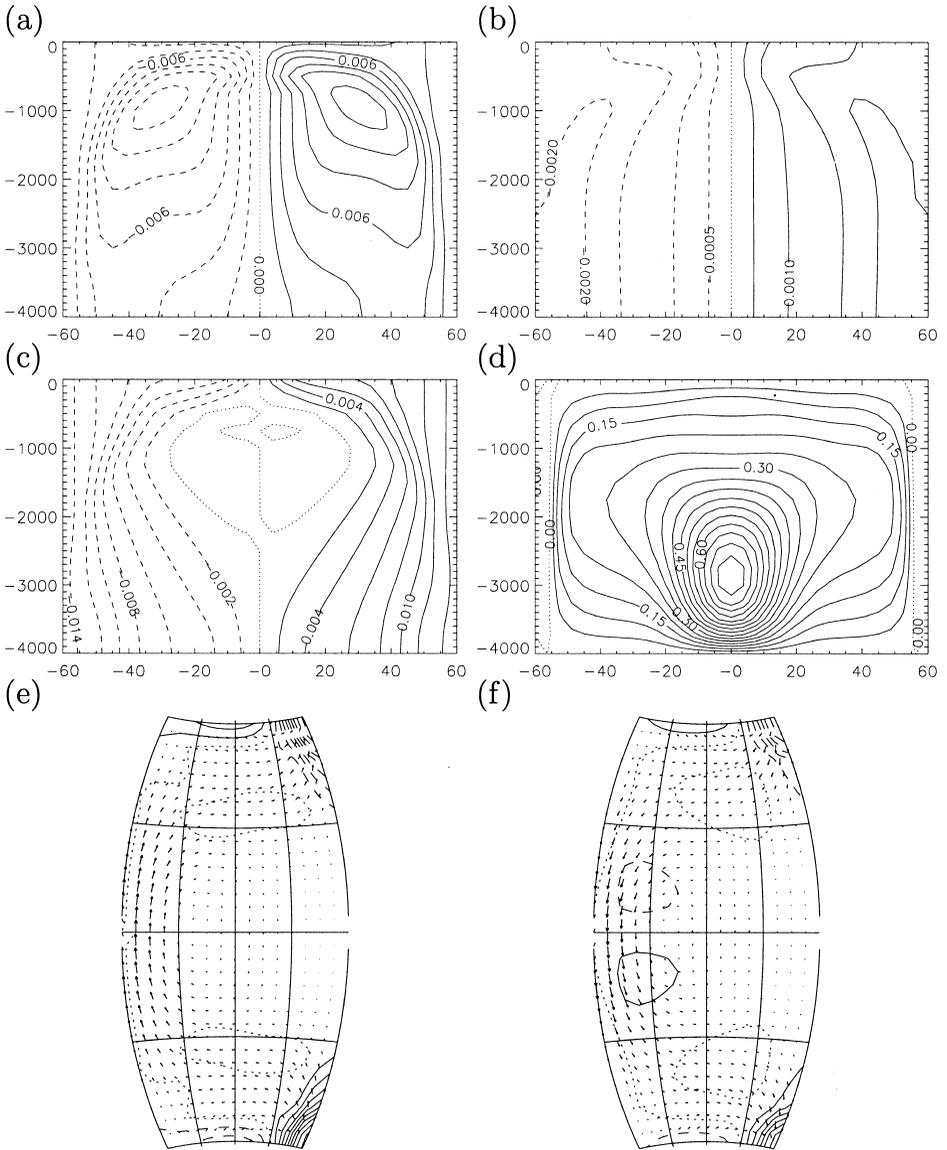
Figure 11. Most unstable eigenvector corresponding to the stably stratified basic state in Figure 10. Same fields as in Figure 7.

nonlinear interaction, and the sign of the energy flow determines whether the perturbation grows or not.

Let us denote the basic solution by $(\bar{\mathbf{u}}, \bar{T}, \bar{S})$ and the destabilizing perturbation at the pitchfork P as $(\mathbf{u}', T', S')$, with $\bar{\rho} = \lambda\bar{S} - \bar{T}$ and $\rho' = \lambda S' - T'$. Volume integration will be denoted by brackets, so that $\langle \cdot \rangle = \int \cdot \, dV$. We define an energy functional through

$$\mathcal{E} = \tfrac{1}{2}\epsilon_R\langle\mathbf{u}'\cdot\mathbf{u}'\rangle + \tfrac{1}{2}Ra\langle(\rho')^2\rangle \tag{17}$$

which is the sum of the kinetic energy of the perturbation and the (squared) norm of the anomalous density field. The latter term resembles the (linearized form of) Available Potential Energy, as used by Bryan and Lewis (1979). The condition for linear instability of the basic state is thus $d\mathcal{E}/dt > 0$. Since inertia does not play a role in this system where $\epsilon_R = 0$, the sources and sinks of kinetic energy instantaneously match, and the perturbation cannot grow by extracting kinetic energy or momentum from the basic state. The time-derivative of $\mathcal{E}$ can be determined using:

$$\frac{1}{2}\frac{d}{dt}\langle(\rho')^2\rangle = \left\langle\rho'\,\frac{\partial\rho'}{\partial t}\right\rangle = \left\langle\rho'\left(\lambda\,\frac{\partial S'}{\partial t} - \frac{\partial T'}{\partial t}\right)\right\rangle \tag{18}$$

where the time-evolution equations for the salinity and temperature anomalies follow from inserting $\bar{T} + T'$ and $\bar{S} + S'$ (and $\bar{\mathbf{u}} + \mathbf{u}'$) in Eqs. (3e) and (3f). After some manipulations, and absorbing $1/Ra$ in $\mathcal{E}$, the time-derivative of $\mathcal{E}$ can be shown to satisfy:

$$\frac{d\mathcal{E}}{dt} = -\langle\rho'\mathbf{u}'\cdot\nabla\bar{\rho}\rangle + B\langle\rho'T'g(z)\rangle - \langle P_H(\nabla_H\rho')^2\rangle - \left\langle P_V\left(\frac{\partial\rho'}{\partial z}\right)^2\right\rangle = P_A + P_B + D_H + D_V \tag{19}$$

According to this expression, the initial growth (or decay) of the anomalous density field ρ' is governed by interaction between the velocity anomaly and the background density distribution (P_A), by interaction with the atmosphere (P_B), and by mixing processes in the ocean, which are parameterized as diffusion (D_H and D_V). The terms in Eq. (19) were evaluated for the most unstable eigenvectors at the pitchfork bifurcations P in Figures 2 and 6. The term P_A was evaluated for the interaction between each particular eigenvector and the basic states at ($\sigma = \sigma(P)$) and surrounding ($\sigma = \sigma(P) \pm \Delta\sigma$) the bifurcation. Although the most unstable eigenvector *at* the pitchfork bifurcation may be suboptimal for $\sigma = \sigma(P) \pm \Delta\sigma$, the changes in the eigenvector will be small compared to changes in the basic states, which respond directly to the changed forcing.

Figure 12 shows that the pitchfork bifurcations P in Figures 2 and 6 are indeed associated with zero-crossings of $d\mathcal{E}/dt$. The instability appears to be strongly promoted by the surface boundary conditions, P_B: the positive thermal anomaly on the southern hemisphere (Fig. 8a) enhances the surface heat loss, and amplifies the positive (salinity dominated) density anomaly (Fig. 8c) on that hemisphere. Note that when restoring conditions are applied on both T and S (say, with relaxation constant B), the term P_B in Eq. (19) becomes $-B\langle\rho'(\lambda S' - T')g(z)\rangle = -B\langle(\rho')^2g(z)\rangle$, which is negative definite. This analysis thus readily demonstrates why a TH circulation, which is stable under restoring boundary conditions, may become unstable when a switch to mixed boundary conditions is made (e.g. Bryan (1986)); the stabilizing influence of restoring conditions on both T and S ($P_B < 0$) at once becomes destabilizing (with $P_B > 0$) when the restoring condition on S is removed.
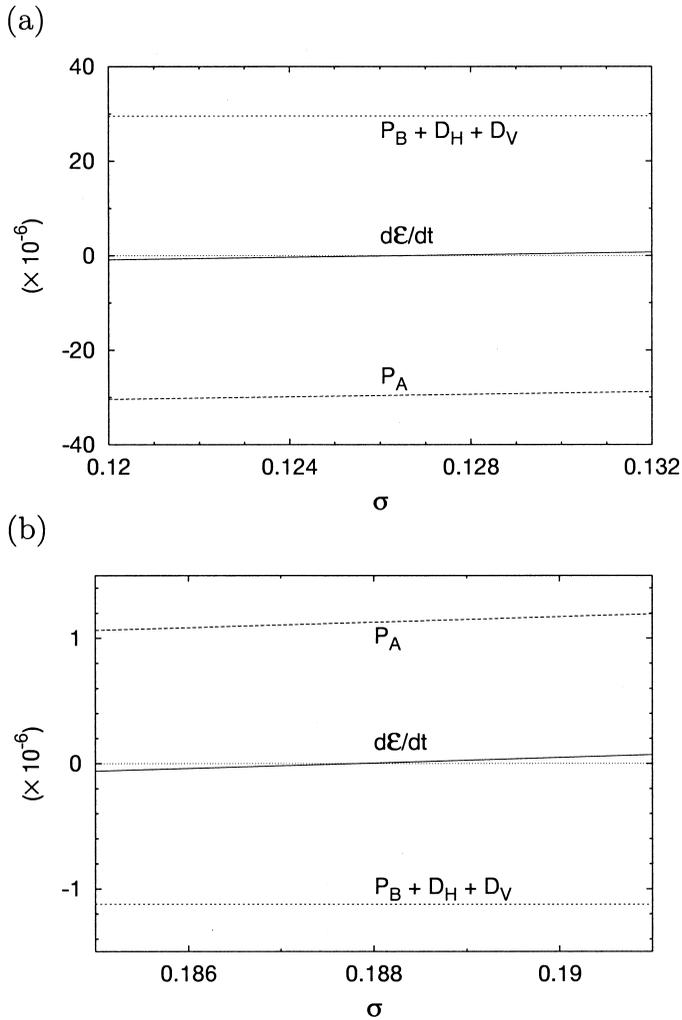
(a)

(b)

Figure 12. The term $d\mathcal{E}/dt$ and its components (Eq. 19) for the most unstable eigenvector at the pitchfork bifurcation P and the surrounding basic states; (a) for the rotationless case of Figure 2 and (b) for the rotational case of Figure 6. The plot shows that the pitchfork bifurcations are associated with zero-crossings of $d\mathcal{E}/dt$. For $d\mathcal{E}/dt > 0$ the most unstable perturbation will grow, rendering the basic TH state unstable. Note that the contributions of the boundary conditions P_B ((a): 66, (b): 4.6), horizontal diffusion D_H ((a): -17, (b): -3.7) and vertical diffusion D_V ((a): -19, (b): -2.0) are relatively large and summed up for convenience. Contrary to the advective term P_A, which also depends on the (changing) background state, these terms only depend on the specific perturbation and are constant. Note that the plots cannot be compared quantitatively, since the amplitudes of the eigenvectors are undetermined.

The perturbation is damped by the horizontal and vertical diffusion terms (D_H, $D_V <$ 0), which represent turbulent mixing processes in the ocean. However, given the perturbation, P_B, D_H and D_V are fixed and cannot be held responsible for the zero-crossing of $d\mathscr{E}/dt$ at the pitchfork bifurcation. It is up to the interaction between the perturbation and the (changing) background stratification (P_A) to determine whether the perturbation will grow or not. This background stratification is determined by both the salinity and temperature fields. The term P_A can therefore be split into a saline component, $P_A^S = -\lambda\langle\rho'\mathbf{u}' \cdot \nabla\bar{S}\rangle$ and a thermal contribution $P_A^T = \langle\rho'\mathbf{u}' \cdot \nabla\bar{T}\rangle$. Each term furthermore consists of zonal, meridional and vertical components, given by $P_{A,x}^S = -\lambda\langle\rho'u'\bar{S}_x\rangle$, $P_{A,y}^S = -\lambda\langle\rho'v'\bar{S}_y\rangle$, $P_{A,z}^S = -\lambda\langle\rho'w'\bar{S}_z\rangle$, etc. Identifying the feedback that is responsible for the increase in P_A with σ, and thus ultimately for the instability, is not straightforward. Figure 13a shows the two feedbacks that contribute to P_A in the rotationless case (Figs. 2 and 12a). The arrow labeled M_a points to the components of the meridional advective feedback, ($P_{A,y}^T$, $P_{A,y}^S$), whose total amplitude is given by $P_{A,y} = P_{A,y}^T + P_{A,y}^S$. This feedback is negative for heat and positive for salt, since the positive density anomaly in the southern hemisphere is attenuated by (anomalous) southward advection of heat from the tropical regions, whereas it is enhanced by advection of salt from the evaporative (sub)tropics. The arrow indicated by V_a indicates the (much weaker) vertical advective feedback. This feedback is positive for temperature and negative for salt, and it is mainly generated in the unstably stratified polar regions. Here, cold (fresh) surface water is advected downward and enhances (attenuates) the positive density anomaly at depth.

Although the sum of the advective feedbacks (P_A) is negative in this rotationless case (see also Fig. 12a), and the perturbation is thus basically damped by the interaction with the background state (mainly due to the meridional temperature advection feedback), the strength of this damping is reduced when σ increases from slightly below to slightly above the pitchfork bifurcation. Figure 13b shows that the tendency of P_A at the pitchfork P is indeed positive when σ is increased. The direction and amplitude of the arrow labeled M_a in Figure 13b indicates that the increase in the meridional salt advection feedback is mainly responsible for the positive tendency of P_A. The increase in the surface salt flux amplitude σ enhances the meridional salt gradient at the surface, and strengthens the meridional salt advection feedback. This tendency is weakly counteracted by a decrease in the (positive) vertical advective feedback (V_a). This decrease reflects the cooling and freshening of the polar surface waters, resulting from a reduced overturning and an increased freshwater flux. This energy analysis is consistent with the mechanistic view of earlier work (Marotzke *et al.,* 1988; Quon and Ghil, 1992; Vellinga, 1996; Dijkstra and Molemaker, 1997) that the meridional salt advection feedback is the main mechanism responsible for the symmetry-breaking bifurcation of the two-dimensional TH state. The situation changes when rotation is added. Figure 14a shows the interaction terms between the destabilizing perturbation and the background state at the pitchfork bifurcation P in Figure 6. Clearly, the importance of the meridional feedback is strongly reduced with respect to the vertical feedback, despite the fact that the addition of rotation and the applied rescaling of Ra
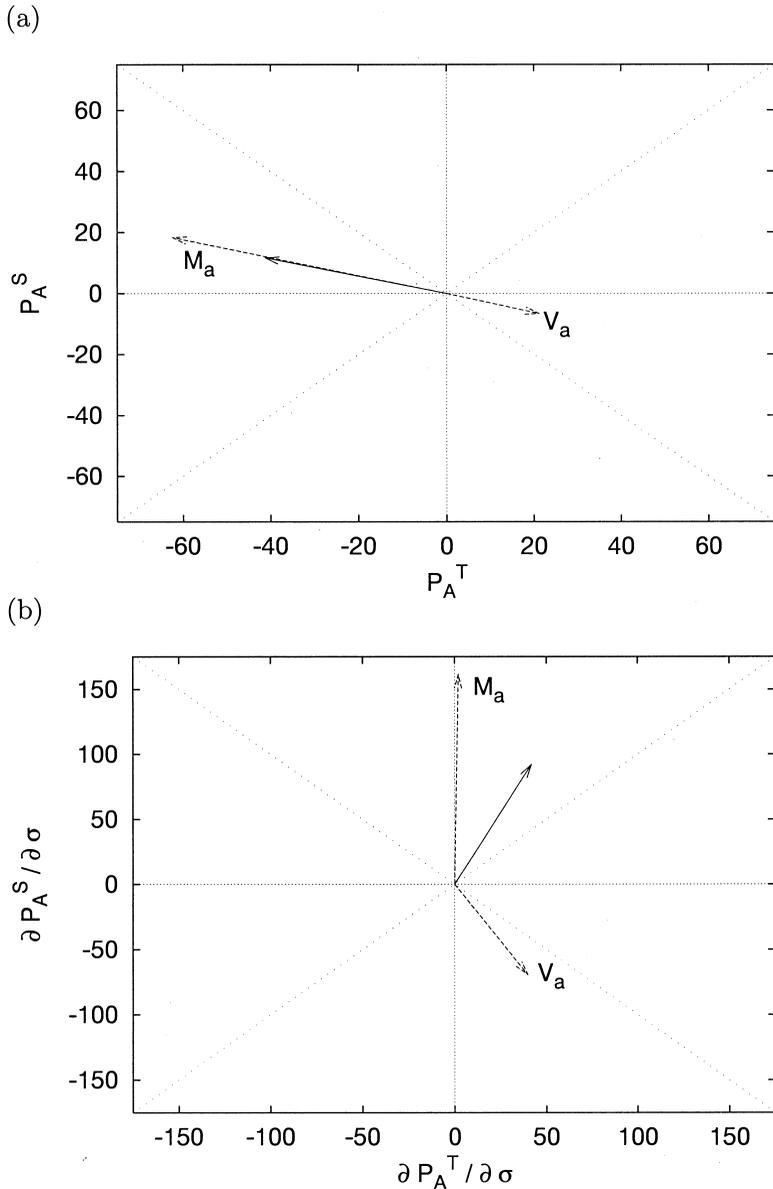
(a)

(b)

Figure 13. (a) The terms P_A^S and P_A^T that make up the advective feedback P_A and (b) their tendencies with changing σ for the rotationless case. The saline contributions are set out against the thermal contributions (solid arrows), and a subdivision is made into the meridional and vertical feedbacks (dashed arrows denoted by M_a and V_a, respectively). The arrows point at the scalar values of the terms corresponding to the interaction between the basic state and its most unstable eigenvector at the pitchfork bifurcation P in Figure 2. The total amplitude of a feedback (or its tendency) is given by $(\partial/\partial\sigma)P_A = (\partial/\partial\sigma)P_A^T + (\partial/\partial\sigma)P_A^S$ (NB: this is not proportional to the length of the corresponding arrow). Points above the line $y = -x$ have a positive contribution to $(\partial/\partial\sigma)P_A$, whereas points below this line have a negative contribution. For points in the regions where $y < |x|$ the thermal contributions dominate, whereas otherwise the saline components are dominant.
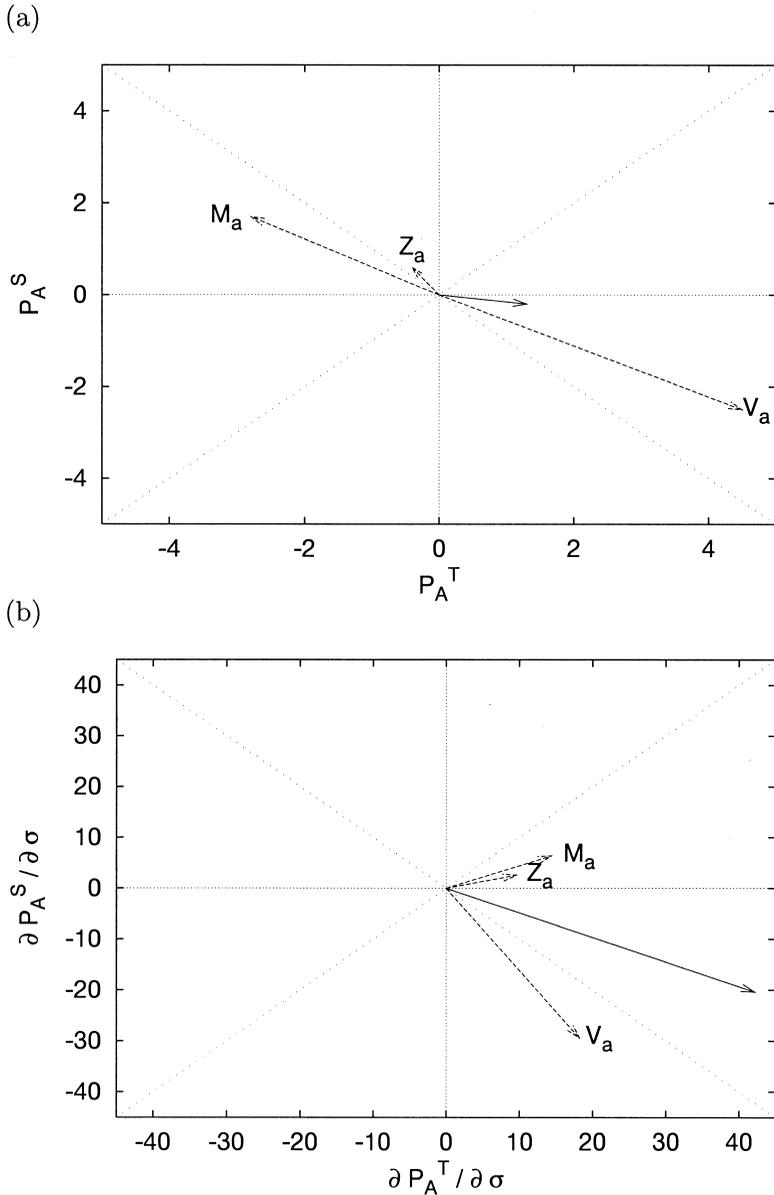
(a)

(b)

Figure 14. Same as Figure 13 but now for the rotational case, i.e. for the basic state (Fig. 7) and destabilizing eigenvector (Fig. 8) at the pitchfork bifurcation P in Figure 6. Rotation breaks the zonal symmetry of the system and generates zonal structure in the eigenvector and the basic states, giving rise to a (weak) zonal advective feedback (denoted Z_a). The meridional salt advection feedback (M_a) is reduced in strength with respect to the vertical advective feedback (V_a). The destabilizing tendencies of several feedbacks are responsible for the zero-crossing of $d\mathscr{C}/dt$.

increased the basic overturning considerably (Fig. 4). However, under influence of rotation and the increased overturning, the main downwelling is forced to take place in narrow boundary layers close to the southern boundary (Figs. 7d–f), rather than in the broad downwelling areas that characterized the rotationless case (Fig. 3a). This enhances the importance of the downwelling in the unstably stratified regions, and strengthens the vertical advective feedback.

The tendency of P_A with increasing σ at the pitchfork is positive, as shown in Figure 14b (and Fig. 12b). However, it is not dominated by an increase in the meridional salt advection feedback, as was the case in the rotationless situation (Fig. 13b). Instead, it reflects destabilizing changes in all but one of the feedbacks; only the vertical salt advection feedback strengthens its stabilizing influence. Thus, in the rotational case, the changes in the background state across the pitchfork bifurcation are complex and result in several destabilizing changes in the advective feedbacks. There is not one advective mechanism responsible for generating the instability at P. The dominant role that is played by the vertical advective feedback when rotation is included is remarkable. It clearly originates in the unstably stratified polar regions, and it thus may be expected to lose its importance when the static instabilities are removed. Therefore, we calculated the same terms as in Figure 14, but now for the eigenvector corresponding to the stably stratified basic states (Fig. 11). Figure 15a shows that the meridional feedback has re-established its role as a dominant feedback mechanism. The vertical feedbacks have strongly decreased in strength, and have reversed their sign. Now the stable (unstable) thermal (saline) stratification in the extra-polar regions provide for a negative (positive) feedback.

Upon changing σ, the increase of the meridional salt advection feedback is again to a large extent responsible for the positive tendency of P_A (Fig. 15b). This feedback has benefitted from the strengthening of the mean meridional salinity gradient by the enhanced surface salt flux. An increase in the mean meridional temperature gradient is responsible for the negative tendency of $P_{A,y}^T$. Nonetheless, the role of the vertical advective feedback has not completely vanished. Its stabilizing influence has diminished by a decrease in the negative heat advection feedback. Inspection of the thermal fields shows that the increase in σ has increased the temperatures at depth in the extra-polar regions. The resulting decrease in the background (positive) vertical temperature gradient means a reduction in the negative feedback.

5. Discussion

The value of two-dimensional or zonally averaged models of the thermohaline circulation has often been questioned. The main concern is the absence (or poor parameterization) of Coriolis effects on these flows. Indeed, *a priori* it would seem that by poorly modeling (or not even representing at all) the dominant geostrophic balance, many correspondences of the two-dimensional and three-dimensional flows would disappear. On the other hand, for the double-hemispheric configuration, many results in low-resolution GCMs (Bryan, 1986; Klinger and Marotzke, 1999) are easily interpreted in terms of those of two-
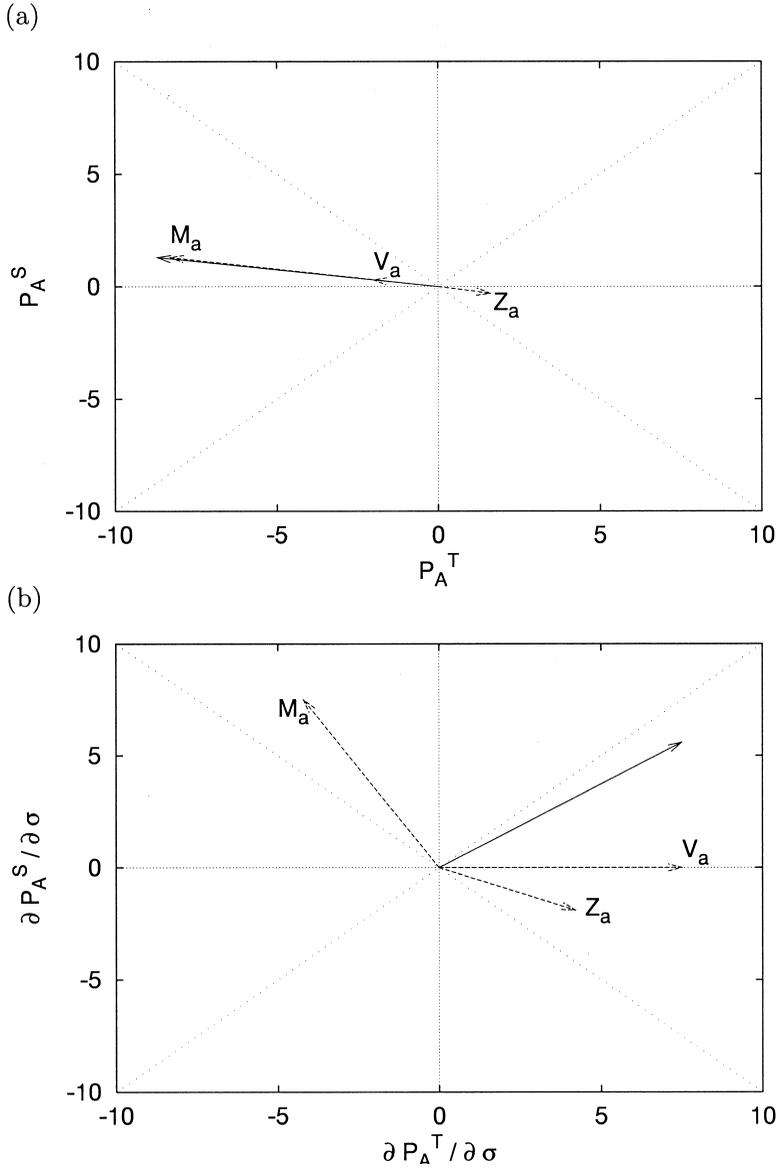
(a)

(b)

Figure 15. Same as Figure 14, but now for the associated stably stratified basic state (Fig. 10) and eigenvector (Fig. 11). Note that the meridional advective feedbacks have re-established their dominant role at the expense of the vertical advective feedbacks. Nonetheless, the decrease of the (negative) vertical heat advection feedback does play a role in destabilizing the basic state.

dimensional models. In particular, the 'guessed' bifurcation diagram in Klinger and Marotzke (1999) is qualitatively very similar to that in Dijkstra and Molemaker (1997). Using the present results, a more definite statement can be made on the relation between steady states in two- and three-dimensional models. Note that this correspondence is restricted to the steady-state structure and not immediately to transient flows; in three-dimensional models, oscillatory instabilities are found (Chen and Ghil, 1995; Te Raa and Dijkstra, 2001), which have no counterpart in two-dimensional models. Furthermore, other purely three-dimensional features may influence the bifurcation structure considerably, as has been shown here for the wind-driven circulation.

Bifurcation studies on two-dimensional equatorially symmetric models, whether in the rotationless limit (Dijkstra and Molemaker, 1997; Quon and Ghil, 1995; Vellinga, 1996; Weijer *et al.,* 1999), or in zonally averaged formulation (Stocker and Wright, 1991; Vellinga, 1996), all show that multiple equilibria arise through a pitchfork bifurcation located on the TH branch (Dijkstra, 2000). A similar pitchfork bifurcation is found here in the three-dimensional model, and it turned out to be robust with respect to wind forcing and convective adjustment. Whether the pitchfork bifurcation is sub- or super-critical depends on the model configuration and parameters. In two-dimensional models, both types of bifurcations are found (Dijkstra and Molemaker, 1997), whereas in the three-dimensional cases here, only super-critical pitchfork bifurcations are found. In Klinger and Marotzke (1999), a subcritical pitchfork was suggested, but the bifurcation diagram in Figure 9 shows how one easily can get fooled by only computing stable steady states through transient integration. By following the NPP/SPP branch for smaller values of σ, a transition to the TH state will occur below values corresponding to L_2. Following the TH branch with increasing σ then leads to an estimate of the location of the pitchfork bifurcation P. However, one would be very lucky to find a solution on the stable branch between P and L_1 and one would tend to conclude (based on the relative location of L_2 and P) that the pitchfork is sub-critical, whereas detailed bifurcation analysis shows that it is super-critical.

Whereas the type of pitchfork bifurcation may be considered a minor issue, the main point of discussion is: are the physical processes responsible for this symmetry-breaking the 'same' in all these models? In each case, instability sets in when a critical value of the freshwater flux strength σ is exceeded. This means that amplification of the perturbation density occurs only when the basic state salinity gradient is sufficiently strong. Moreover, the zonally averaged patterns of the mode destabilizing the TH state (determined from the eigenvector at the pitchfork bifurcation) correspond remarkably well. They are characterized by antisymmetric temperature and salinity perturbations. The salinity perturbation dominates the density perturbation, which is equatorially antisymmetric as well.

At this point, differences between rotationless and rotational models arise. In two-dimensional Boussinesq models such as antisymmetric density perturbation forces a direct frictional-buoyancy-driven perturbation flow. In the three-dimensional case, it sets up a mainly zonal perturbation flow; a meridional flow is only generated indirectly through a

perturbation east-west pressure gradient. The spatial patterns of the perturbation flows thus obviously differ between rotationless and rotational cases. Also, as the structure of the basic states is quite different between these models, and the values of $\sigma(P)$ differ quantitatively, the importance of the salt advection feedback in each case is not very clear; it is difficult to assess whether the mechanisms are indeed the 'same.'

This notion can be made more precise by characterization of the pitchfork bifurcation through energy analysis, using the norm of the density perturbations $\mathcal{E}$ as energy functional. Indeed, the location of the pitchfork bifurcations in parameter space is in both two- and three-dimensional cases associated with a zero of $d\mathcal{E}/dt$. This zero is determined by the strength of the diffusive damping of the density perturbation, and by possible amplification mechanisms like energy exchange with the basic state (P_A) or surface forcing (P_B). The present analysis indicates that the restoring condition on surface temperature is an important factor destabilizing the TH state ($P_B > 0$). It readily demonstrates why a TH circulation, that is stable under restoring boundary conditions, may become unstable when a switch to mixed boundary conditions is made (e.g. Bryan (1986)): restoring conditions on both T and S make the term P_B in Eq. (19) negative, and thus strongly stabilize the circulation. Upon switching to mixed boundary conditions, P_B becomes positive, and the boundary conditions tend to destabilize the flow.

The most relevant energy-producing terms are associated with spatial correlations between the density anomalies that drive the perturbation flow, and those which are generated by this perturbation flow, via anomalous advection of basic-state temperature and salinity. The volume integrated values of these production terms, P_A^T and P_A^S, largely determine the location of the critical value of σ. The results for the two-dimensional case show that the term $P_{A,y}^S$, associated with the meridional salt advection feedback, is the dominant destabilizing term. This feedback has been shown to play an important role in the stably stratified three-dimensional case too. However, its identification as *the* destabilizing mechanism is not fully justified here; the increase in σ affects the basic-state temperature and salinity fields to such an extent, that other feedbacks display destabilizing tendencies as well.

There are subtleties in the production of energy when an unstable stratification of the basic state is present. Small regions of unstable stratification in the background flow may lead to an additional energy production, which in principle destabilizes the symmetric TH flow. This may explain the result found in Marotzke (1991), that the flows using the classical adjustment scheme in the MOM model are more unstable than those using a scheme which guarantees stably stratified solutions. However, convective adjustment is shown here to perform two tasks at the same time: by removing static instabilities these 'artificial' energy production terms are reduced to near zero, but by modifying the global pressure distribution the overturning strength of the basic state increases, enhancing the energy production term associated with meridional salt advection. In the model configuration here, the net effect is a more destabilized TH state, but these effects may be quite model-dependent.

Although details of the instability mechanism determine the exact location of the pitchfork bifurcation in parameter space and may be quite model-dependent, the meridional salt advection feedback may be regarded as the destabilizing mechanism that is central in two- and three-dimensional models; it is the only feedback that directly depends on the surface salt flux amplitude σ. We thus conclude that the physical mechanism of symmetry-breaking is essentially the same in both two- and three-dimensional models. It indeed appears that the details of the momentum balances are not central to this symmetry-breaking mechanism as long as there is an overturning response to a meridional density gradient.

APPENDIX

A. Continuation of steady states and their stability

The spatially discretized model equations can be written in the form

$$\mathbf{M} \frac{d\mathbf{u}}{dt} = \mathbf{F}(\mathbf{u}) = \mathbf{L}(\mathbf{u}) + \mathbf{N}(\mathbf{u}, \mathbf{u}) \tag{A1}$$

where the vector $\mathbf{u}$ contains the unknowns (u, v, w, p, T, S) at each grid point and hence has dimension $d = 6 \times N \times M \times L$. The operators $\mathbf{M}$ and $\mathbf{L}$ are linear and $\mathbf{N}$ represents the nonlinear terms in the equations. Steady state solutions lead to a set of nonlinear algebraic equations of the form

$$\mathbf{F}(\mathbf{u}, \mathbf{p}) = 0 \tag{A2}$$

where the parameter dependence of the equations is made explicit through the p-dimensional vector of parameters $\mathbf{p}$ and hence $\mathbf{F}$ is a nonlinear mapping from $\mathbf{R}^{d+p} \to \mathbf{R}^d$. As can be readily seen from the continuous form of the steady equations, the salinity is determined up to an additive constant. Moreover, also the pressure is determined up to an additive constant. To calculate a steady-state solution of the system of equations, the equations are regularized (such that the Jacobian matrix is nonsingular) by fixing the pressure at a particular point (in our case at the point (N, M, L)). In addition, an integral condition for S is substituted for the last equation from the salinity equation, such that salt is conserved exactly within the domain. Since the total dimensional salt content is $\rho_0 V S_0$, where V is the total volume of the basin, the scaling for salinity provides the dimensionless form as

$$\int_V S \cos \phi \, d\lambda \, d\phi \, dz = 0 \tag{A3}$$

which is a constraint on the deviation of the salinity field from uniform conditions. To determine branches of steady solutions of the equations (A2, A3) as one of the parameters, say μ, is varied, the pseudo-arclength method (Keller, 1977) is used. The branches $(\mathbf{u}(s), \mu(s))$ are parameterized by an 'arclength' parameter s. An additional equation is obtained by 'normalizing' the tangent

$$\dot{\mathbf{u}}_0^T(\mathbf{u} - \mathbf{u}_0)\dot{\mu}_0(\mu - \mu_0) - \Delta s = 0 \tag{A4}$$

where $(\mathbf{u}_0, \mu_0)$ is an analytically known starting solution or a previously computed point on a particular branch and Δs is the step-length. To solve the system of equations (A2–A4), the Newton-Raphson method is used. The linear systems are solved with the BICGSTAB method using the MRILU preconditioning technique (Dijkstra *et al.*, 2001).

When a steady state is determined, the linear stability of the solution is considered and transitions that mark qualitative changes, such as transitions to multiple equilibria (pitchfork or saddle node bifurcations) or periodic behavior (Hopf bifurcations), can be detected. The linear stability analysis amounts to solving a generalized eigenvalue problem of the form

$$\alpha \mathcal{A}\mathbf{x} = \beta \mathcal{B}\mathbf{x} \tag{A5}$$

where $\mathcal{A}$ is the Jacobian matrix (the derivative $\mathbf{F}_u$) and $\mathcal{B} = -\mathbf{M}$, which are in general nonsymmetric matrices. If $\mathcal{B}$ is nonsingular, the problem reduces to an ordinary eigenvalue problem for the matrix $\mathcal{B}^{-1}\mathcal{A}$. Because only real matrices are considered, there are d eigenvalues which are either real or occur as complex conjugate pairs. However, if $\mathcal{B}$ is singular, the eigenvalue structure may be more complicated; the set of eigenvalues may be finite, empty or even the whole complex plane (Golub and Van Loan, 1983). In the particular model here, $\mathcal{B}$ is a singular diagonal matrix because time derivatives are absent in the continuity equation and vertical momentum equation. The problem (A5) is solved by the Jacobi-Davidson QZ-method (JDQZ). With this method, one can compute several, say m, eigenvalues and optionally eigenvectors near a specified target τ. The details of the method are described elsewhere (Sleijpen and Van der Vorst, 1996) and the implementation of JDQZ in an earlier version of our continuation code in Van Dorsselaer (1997). For $\alpha \neq 0$, we write $\mathbf{s}_r + i\mathbf{s}_i = \beta/\alpha$ for the real and imaginary parts of the eigenvalue.

B. The Global Adjustment Procedure

The Global Adjustment Procedure (GAP) starts off with $\mathbf{u}^1 = \mathbf{u}$, and a constant field of vertical diffusivity $P_V^1(\lambda, \phi, z) = P_V^0$, with P_V^0 the standard value of vertical diffusivity as in Table 1. Within a step k of an iterative loop over N_a steps, a stably stratified solution $\mathbf{u}_{st}^k$ is constructed from $\mathbf{u}^k$, using the convective adjustment procedure of Rahmstorf (1993). A linear combination is taken:

$$\tilde{\mathbf{u}}^k = (1 - \omega_k)\mathbf{u}^k + \omega_k\mathbf{u}_{st}^k \qquad k = 1, N_a \tag{B1}$$

where ω_k increases from zero to unity in N_a steps (according to $\omega_k = \frac{1}{2}(1 - \cos \pi k/N_a)$). The vertical diffusivities are adjusted according to

$$P_V^{k+1} = P_V^c \tanh \left(\Gamma \frac{P_V^k}{P_V^c} \right); \Gamma \equiv \frac{\partial \rho^k/\partial z}{\partial \tilde{\rho}^k/\partial z} \qquad (B2)$$

where P_V^c is an upper bound on the vertical diffusivities. In regions that are stabilized by a pass of the adjustment procedure, $\Gamma > 1$. For modest changes in the stratification (e.g., when ω_k is still small), this procedure guarantees that the vertical diffusive fluxes of buoyancy associated with $\tilde{\mathbf{u}}^k$ and $\mathbf{u}^k$ are the same. For large values of Γ (i.e., in well mixed areas when $\omega_k \rightarrow 1$), the vertical diffusivity is bounded by P_V^c. A Newton step is performed on $\tilde{\mathbf{u}}^k$ to obtain a new estimate $\mathbf{u}^{k+1}$, which includes an update of the velocity field. After N_a steps, the procedure is repeated with $\omega_k = 1$ for $k > N_a$ until convergence is reached. The resulting solution $\mathbf{u}_{st}$ is stably stratified, has enhanced diffusivities in the regions where convection took place, and has a velocity field that is consistent with the density field.

REFERENCES

Bradley, R. S. 1999. Paleoclimatology: Reconstructing Climates of the Quaternary, Harcourt/Academic Press, San Diego, CA, 613 pp.

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

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

Bryan, K. and L. J. Lewis. 1979. A water mass model of the world ocean. J. Geophys. Res., *84,* 2503–2517.

Cessi, P. and W. R. Young. 1992. Multiple equilibria in two-dimensional thermohaline circulation. J. Fluid Mech., *241,* 291–309.

Chen, F. and M. Ghil. 1995. Interdecadal variability of the thermohaline circulation and high-latitude surface fluxes. J. Phys. Oceanogr., *22,* 161–167.

Dijkstra, H. A. 2000. Nonlinear Physical Oceanography, Kluwer Academic Publishers, Dordrecht, The Netherlands, 456 pp.

Dijkstra, H. A. and M. J. Molemaker. 1997. Symmetry breaking and overturning oscillations in thermohaline-driven flows. J. Fluid Mech., *331,* 195–232.

Dijkstra, H. A., H. Oksuzoglu, F. W. Wubs and E. F. F. Botta. 2001. A fully implicit model of the three-dimensional thermohaline ocean circulation. J. Comp. Physics, (submitted).

Golub, G. H. and C. F. Van Loan. 1983. Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 642 pp.

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

Keller, H. B. 1977. Numerical solution of bifurcation and nonlinear eigenvalue problems, *in* Applications of Bifurcation Theory, P. H. Rabinowitz, ed., Academic Press, NY, 359–389.

Killworth, P. D. 1985. A two-level wind and buoyancy driven thermocline model. J. Phys. Oceanogr., *15,* 1414–1432.

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

Marotzke, J. 1991. Influence of convective adjustment on the stability of the thermohaline circulation. J. Phys. Oceanogr., *21,* 903–907.

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

Quon, C. and M. Ghil. 1992. Multiple equilibria in thermosolutal convection due to salt-flux boundary conditions. J. Fluid Mech., *245,* 449–484.

——— 1995. Multiple equilibria and stable oscillations in thermosolutal convection at small aspect ratio. J. Fluid Mech., *291,* 33–56.

Rahmstorf, S. 1993. A fast and complete convection scheme for ocean models. Ocean Model., *101,* 9–11.

Sleijpen, G. L. G. and H. A. Van der Vorst. 1996. A Jacobi-Davidson iteration method for linear eigenvalue problems. SIAM J. Matrix Anal. Appl., *17,* 410–425.

Stocker, T. F. 2000. Past and future reorganizations in the climate system. Quater. Sci. Rev., *19,* 301–319.

Stocker, T. F. and D. G. Wright. 1991. A zonally averaged ocean model for the thermohaline circulation. II: Interocean circulation in the Pacific–Atlantic Basin System. J. Phys. Oceanogr., *21,* 1725–1739.

Stommel, H. 1961. Thermohaline convection with two stable regimes of flow. Tellus, *2,* 244–230.

Te Raa, L. and H. A. Dijkstra. 2001. Instability of the thermohaline ocean circulation on interdecadal time scales. J. Phys. Oceanogr. (in press).

Thual, O. and J. C. McWilliams. 1992. The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models. Geophys. Astrophys. Fluid Dyn., *64,* 67–95.

Tziperman, E. 2000. Proximity of the present-day thermohaline circulation to an instability threshold. J. Phys. Oceanogr., *30,* 90–104.

Van Dorsselaer, J. J. 1997. Computing eigenvalues occurring in continuation methods with the Jacobi-Davidson QZ method. J. Comp. Physics, *138,* 714–733.

Vellinga, M. 1996. Instability of two-dimensional thermohaline circulation. J. Phys. Oceanogr., *26,* 305–319.

——— 1998. Multiple equilibria of the thermohaline circulation as a side effect of convective adjustment. J. Phys. Oceanogr., *28,* 305–319.

Weaver, A. J. and T. M. Hughes. 1996. On the incompatibility of ocean and atmosphere and the need for flux adjustments. Climate Dyn., *12,* 141–170.

Weijer, W., W. P. M. De Ruijter, H. A. Dijkstra and P. J. Van Leeuwen. 1999. Impact of interbasin exchange on Atlantic overturning. J. Phys. Oceanogr., *29,* 2266–2284.

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

Winton, M. 1996. The role of horizontal boundaries in parameter sensitivity and decadal-scale variability of coarse-resolution ocean general circulation models. J. Phys. Oceanogr., *26,* 289–304.

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.