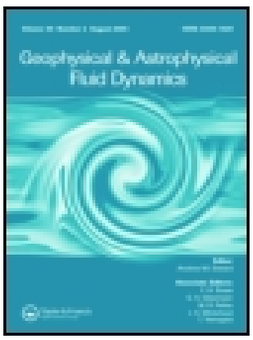

Geophysical & Astrophysical Fluid Dynamics

ISSN: 0309-1929 (Print) 1029-0419 (Online) Journal homepage: https://www.tandfonline.com/loi/ggaf20

The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models

Olivier Thual & James C. Mcwilliams

Geophys. Astrophys. Fluid Dynamics, Vol. 64, pp. 67–95
Reprints available directly from the publisher
Photocopying permitted by license only

THE CATASTROPHE STRUCTURE OF THERMOHALINE CONVECTION IN A TWO-DIMENSIONAL FLUID MODEL AND A COMPARISON WITH LOW-ORDER BOX MODELS

OLIVIER THUAL and JAMES C. McWILLIAMS

NCAR, PO Box 3000, Boulder, CO 80307, USA

(*Received 18 December 1991; in final form 1 August 1991*)

We impose a surface forcing on the 2D, Boussinesq, thermohaline equations in a rectangular domain, in the form of equatorially symmetric cosine distributions of salinity flux and temperature. This system may be seen as an idealization of the ocean thermohaline circulation on the global scale over intervals of centuries or millenia. Multiple steady states are found numerically. They reflect the competition between the opposite signs of the temperature and salinity-driven equatorially symmetric circulations. There are also pole-to-pole, equatorially asymmetric circulations. In the control space of the temperature and salinity-flux forcing amplitudes, these equilibria form two cusp catastrophes, and transitions between stable equilibria occur through several distinct bifurcations. These catastrophes can be reproduced in simple box models connecting stirred reservoirs through capillary pipes. This steady-state analysis may provide a framework for a better understanding of climatic transitions between different stable regimes of the ocean-atmosphere system.

KEY WORDS: Thermohaline convection, ocean circulation, climate models.

1. INTRODUCTION

Several recent studies are concerned with the existence of multiple stable equilibria for the oceanic thermohaline circulation. This phenomenon was first shown by Stommel (1961) with a simple two-box model comprising one equatorial box, which is heated and loses fresh water, and one polar box, which is cooled and receives fresh water (see Figure 5).

Generalizing this approach to a three-box model with north and south polar boxes, Rooth (1982) suggested that, even with symmetric forcing, a symmetric solution could become unstable and develop into a single asymmetric pole-to-pole circulation. Some variants of this box model have been constructed by adding more boxes, considering alternative forcings, or changing the transfer laws between boxes (Walin, 1985; Huang *et al.*, 1990; Birchfield, 1989; Marotzke, 1990). In a review of these box models, Welander (1988) showed that the equilibria of a certain class of three-box models, or even n-box models (see Figure 7), could be viewed as the "superposition" of elementary two-box model equilibria. He showed, for instance, that an asymmetric pole-to-pole circulation was the combination of a temperature-driven two-box cell in one hemisphere, and a salinity-driven two-box cell in the other hemisphere.

* Present address: CERFACS, 42 av. Coriolis, 31057, Toulouse Cedex, France.

Based upon this and geological evidence, Broecker *et al.* (1989), among others, suggested that the ocean-atmosphere system has more than one stable regime. Part of the interest in multiple equilibria for thermohaline circulation comes from its probable role in multiple equilibria for climate.

The existence of multiple equilibria for a 3D oceanic circulation model was demonstrated by Bryan (1986) for an idealized sector geometry and an equatorially symmetric geography and surface forcing. He found two stable steady states: one thermally driven symmetric circulation, with formation of deep water in the polar regions, and one pole-to-pole circulation. The methodology was to make a first calculation with "restoring boundary conditions" (fixed temperature and salinity imposed at the surface) and then diagnose the surface salinity flux of the equilibrium state to impose "mixed boundary conditions" (fixed temperature and fixed salinity flux) in a second calculation, initiated with a perturbation of the equilibrium state. With a similar methodology, multiple equilibria also have been demonstrated in a coupled ocean-atmosphere general circulation model (Manabe *et al.*, 1988).

In a model with complexity intermediate between the box models and the 3D numerical simulations, Marotzke *et al.* (1988) studied 2D thermohaline circulations. After following the methodology of Bryan (1986), they concluded that the thermally driven symmetric circulation is unstable, and that only the pole-to-pole circulation is stable under mixed boundary conditions. When equatorial symmetry is imposed, this model also exhibits a stable salinity-driven symmetric circulation, and its temperature-driven circulation become stable (Marotzke, 1989).

There thus seems to be a conflict between the 3D and 2D solutions concerning the stability of the steady states. To settle this issue, we started by trying to reproduce the Marotzke *et al.* (1988) calculations. It appears that with the choice of zero horizontal diffusivities the solutions become singular in a time integration. We posit that their numerical scheme indeed generates a horizontal diffusivity as an essential element in obtaining non-singular solutions, and we then explore this additional parameter for the problem. We thus count properly in Section 2 the number of parameters in the 2D Boussinesq model. Besides reproducing an example of the restoring and mixed boundary conditions methodology, we focus more on the direct imposition of a cosine surface salinity flux as well as a cosine surface temperature. However, the two procedures can be connected with a simple argument given at the end of Section 3. In Section 3 we perform a systematic exploration of the control space, and come to the conclusion that, depending on the choice of the parameters, the conclusions of both Bryan (1986) and Marotzke *et al.* (1988) can be correct in 2D flow.

To give a global picture of these steady states in the control space, we refer to the theories of catastrophes (Arnold, 1984) and of bifurcations (Arnold, 1989) for dissipative dynamical systems, more for their aptness of language than for all their mathematical power. Indeed, only the simplest catastrophes occur here, the fold catastrophe (for instance the projection of a sphere on a plane) and the cusp catastrophe, which corresponds to the intersection of two folds. The fold is said to be a codimension 1 catastrophe, as one must vary one parameter to cross it. The cusp is of codimension 2, since it is the intersection of two codimension 1 surfaces.

We also deal with the simplest bifurcations of an equilibrium, the saddle-node bifurcation, when two equilibria coalesce, and the pitchfork bifurcation, when an equilibrium invariant under a symmetry (about the equator here) is destabilized and two asymmetric, bifurcated states (here the pole-to-pole circulations) appear that can be transformed into each other by the symmetry; thus, the pitchfork bifurcation is said to break the symmetry. For this bifurcation only two cases are possible: the supercritical case, for which the bifurcated states are stable and are found at the values of the control parameter such that the bifurcating equilibrium is unstable, and the subcritical case, for which the bifurcated states are unstable and coexist with the stable equilibrium. (Here we find only subcritical pitchfork bifurcations.) The saddle-node bifurcation is generic (i.e., among the most likely kinds of bifurcation) in the space of dissipative systems unconstrained by any symmetry, but the pitchfork bifurcation is only generic in systems admitting a symmetry invariance. In our analysis, where we have incomplete dynamical information, we will complete it based on likelihood arguments about generic bifurcation behavior.

We also apply this global view to several box models that have both many precedents (see above) and many fewer degrees of freedom than the fluid model. Of course, the value of a box model comes from its simplicity, but only if its solution behavior can be shown to be apt. We solve analytically the catastrophe structure of the original Stommel two-box model, and it appears to match surprisingly well the catastrophe structure of the equatorially symmetric 2D fluid model (Section 4). We investigate in the same way the three-box model in Section 5 and formulate the "superposition principle" of Welander (1988) in this catastrophe language: in an asymptotic limit, associated with small horizontal diffusivity, the cusp corresponding to the pole-to-pole circulation is identical to the cusp of the symmetric states. In Section 6, we present a hierarchy of box models, within which the two-box and three-box models appear as limit cases. This increase in the number of parameters helps to make a more plausible physical connection with the fluid model. Their solutions also exhibit qualitatively the same catastrophe structure.

2. THE 2D BOUSSINESQ MODEL

We consider a 2D thermohaline model in which the temperature and salinity flux are imposed at the surface. We choose a dimensionless form that is convenient for the asymptotic limits that we will consider later. We indicate how this model can be seen as an idealization of oceanic thermohaline convection.

2.1 Boussinesq equations

We consider a 2D (y, z) layer of fluid in a rectangular basin of width L and depth d. The thermohaline equations are

$$\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} = -\rho_0^{-1}\nabla p + B(T, S)\mathbf{e}_z + \nu\nabla^2\mathbf{u},$$

$$\nabla\cdot\mathbf{u} = 0, \tag{2.1}$$

$$\partial_t T + \mathbf{u}\cdot\nabla T = \kappa_T \nabla^2 T,$$

$$\partial_t S + \mathbf{u}\cdot\nabla S = \kappa_S \nabla^2 S,$$

where $\mathbf{u}$ is the velocity vector field, p is the pressure, ρ_0 is an average value of the density ρ, T and S are the temperature and salinity scalar fields, and $\mathbf{e}_z$ is the vertical unit vector. The dissipation parameters are the viscosity ν and the thermal and saline conductivities κ_T and κ_S. The bouyancy $B(T, S)$ is expressed by the linear equation of state

$$B(T, S) = -g(\rho/\rho_0) = g(\gamma_T T - \gamma_S S), \tag{2.2}$$

where g is the gravitational acceleration, and γ_T and $-\gamma_S$ are the thermal and saline expansion coefficients.

2.2 Non-dimensionalization

In order to write these equations in a dimensionless form, we choose the following units: d for length, $2\pi d^3/L\kappa_T$ for time, $L\kappa_T/2\pi d^2$ for velocity (as a consequence of the previous choices), $\nu\kappa_T L^2/(4\pi^2 d^5 g\gamma_T)$ for temperature, and $\nu\kappa_T L^2/(4\pi^2 d^5 g\gamma_S)$ for salinity. The same units of length and velocity are used in both spatial directions. We also eliminate the pressure by introducing the streamfunction Ψ such that $v = -\Psi_z$ and $w = \Psi_y$ are the horizontal and vertical velocities. With the notation $J(f, g) = f_y g_z - f_z g_y$ for the Jacobian operator, the dimensionless equations are

$$\begin{aligned}
(k\sigma)^{-1}[\partial_t \nabla^2\Psi + J(\Psi, \nabla^2\Psi)] &= k^{-1}(T_y - S_y) + \nabla^4\Psi, \\
k^{-1}[\partial_t T + J(\Psi, T)] &= \nabla^2 T, \\
k^{-1}[\partial_t S + J(\Psi, S)] &= \tau\nabla^2 S,
\end{aligned} \tag{2.3}$$

where $\sigma = \nu/\kappa_T$ is the Prandtl number, $\tau = \kappa_S/\kappa_T$ is the Lewis number, and $k = 2\pi d/L$ is the fundamental wave number.

2.3 Surface forcing and boundary conditions

We force the thermohaline convection by imposing the following surface boundary conditions, expressed in dimensionless form

$$T = a\cos ky \quad \text{and} \quad S_z = b\cos ky, \tag{2.4}$$

at $z = 0$. There is thus a strong discrepancy between the temperature forcing (fixed value) and the salinity forcing (fixed flux). This forcing is equatorially symmetric, since we choose the origin of y at the middle of the domain $-1 \leqslant z \leqslant 0$, $-\pi/k \leqslant y \leqslant \pi/k$.

On the other sides of the domain (bottom and lateral), we choose the no-flux

boundary conditions for the scalar fields

$$\partial_n T = 0 \quad \text{and} \quad \partial_n S = 0, \tag{2.5}$$

where ∂_n is the derivative in the direction perpendicular to the boundary. For the velocity field we choose free-slip boundary conditions on all the sides of the domain:

$$\Psi = 0 \quad \text{and} \quad \partial_n^2 \Psi = 0. \tag{2.6}$$

For the present study, we have compared (2.5)–(2.6) to two variants. In one variant, we choose horizontally periodic conditions (period $2\pi/k$) and check that they give the same results concerning the symmetric solutions. In the other we consider a periodicity of twice the width of the domain ($4\pi/k$), and observe that the symmetric and asymmetric solutions are nearly the same (with nearly the same stability properties) as for the free-slip and no-flux lateral boundary conditions in the closed domain of width ($2\pi/k$). After these checks we have always used one of these two variants for numerical efficiency.

Thus, our 2D thermohaline Boussinesq model is controlled by five dimensionless parameters, (a, b, k, σ, τ). In these coordinates of the control space, we can consider the infinite Prandtl number limit $\sigma \to \infty$, for which the vorticity equation is simply

$$0 = k^{-1}(T_y - S_y) + \nabla^4 \Psi. \tag{2.7}$$

We have checked, for the present study, that there is only a modest difference (at most 5%) between the limit and the case $\sigma = 1$, for the steady states of the regions of the parameters that we have explored. Further exploration of this parameter, in particular in the limit $\sigma \to 0$, could be interesting but is not undertaken here. We have also made the particular choice of $\tau = 1$ to reduce the control space to the parameters (a, b, k).

2.4 Oceanic pertinence of the model

For the idealized 2D Boussinesq problem just posed, one may ask what its relevance is for 3D oceanic thermohaline convection, and the answer we give here is more heuristic than rigorous. First, the 2D fluid can be thought of as a zonally averaged equation of motion, relevant for the longtime scale (of order thousand years), global spatial scale, meridional circulation. Because of this averaging and due to the existence of zonal boundaries of oceans, the zonal velocity is approximately zero and the Coriolis force can be neglected. The viscosity ν and conductivities κ_T and κ_S in the model must be seen as turbulent, or "eddy" diffusivities, representing the transports by smaller scale motions, spanning the scale range from molecular dissipation to the wind-driven gyres. If one believes that the small scale turbulence can be modeled by eddy diffusivities on time scale of the order of months or years, there is no reason in principle why the same parametrization should not be used for longer time scales. On the contrary this parameterization may be even more pertinent on very long time

scales, especially if there is an energy gap in the temporal spectrum. Because of the rotation and the stratification of the ocean, this turbulent viscosity should be anisotropic. However, in the simple 2D model presented above, this anisotropy can be removed by a stretching of the horizontal coordinate e.g.,

$$\nu_v \partial_z^2 + \nu_h \partial_y^2 = \nu_v (\partial_z^2 + \partial_{\tilde{y}}^2), \tag{2.8}$$

where $y = \tilde{y}\sqrt{\nu_h/\nu_v}$. Thus, anisotropy of diffusivities and the aspect ratio of the domain combine in the model parameter k. We assume that the anisotropies of the turbulent scalar conductivities are all equal to the viscosity anisotropy, or that the turbulent Prandtl and Schmidt numbers are independent of direction. Finally, we argue that all scalar properties have roughly equivalent "eddy" transports to justify the choice of $\tau = 1$ in the model. The choice of $\sigma = \infty$ can be rationalized by saying that the nonlinear transport terms of the momentum equation have been parametrized as diffusion, with a turbulent viscosity that is large compared to the diffusivity of material properties. This is perhaps the boldest of our assertions. The insensitivity of our solutions to σ, mentioned in Section 2.3, is our second line of defense.

The choice of mixed boundary conditions in the model is a idealization of the actual forcing of the ocean. The atmosphere acts somewhat like a thermostat which damps any deviation in the surface temperature of the ocean (and vice versa, of course), and its meridional structure is hot at the equator and cold at the poles i.e., $a > 0$ in (2.4). But the evaporation and precipitation processes only imposes a flux of fresh water at the ocean surface (or, equivalently, a salinity flux), without regard to the local value of surface salinity. On the global scale, an excess of fresh water is removed at the equator by evaporation and an excess of precipitation occurs at the pole (Bryan and Oort, 1984), hence $b > 0$ in (2.4).

We do not pretend to reproduce exactly the glaciation/inter-glaciation transitions with this idealized model. Our aim is only to understand and reproduce the basic phenomenon of multiple equilibria and isolate the basic mechanism responsible for it. This simple model is more likely to predict the instabilities of the multiple equilibria than it is to depict accurately the switching behavior between them; so we do not investigate this feature. The patterns of these transitions are likely to be dependant on effects which are ignored in this model, such as the Coriolis force, the topography, the complexity of the surface forcings, etc.

3. THE STATIONARY STATES OF THE FLUID MODEL

The diffusive regime, for small values of the surface forcing, can be computed analytically. For large enough values of the forcing, numerical simulations are required to include nonlinear effects due to advection. Cusp catastrophes (i.e., two folds meeting in a cusp) are found in the nonlinear regime in the space of the two forcing parameters, reflecting the competition between several stable equilibria.

3.1 The diffusive regime

When the surface forcing is small ($a, b \ll 1$), the nonlinear terms can be neglected

in (2.2). Thus T and S are harmonic fields subject to simple boundary conditions, and can be expressed analytically

$$T(y, z) = a \cos ky \frac{\cosh k(z+1)}{\cosh k} \quad \text{and} \quad S(y, z) = b \cos ky \frac{\cosh k(z+1)}{k \sinh k}. \tag{3.1}$$

The associated stream function is

$$\Psi(y, z) = [b - ak \tanh k] \sin(ky)\phi(kz)/k^4, \tag{3.2}$$

where $\phi(Z)$ is given in Appendix A. Contours of the temperature, salinity and streamfunction fields are displayed in Figure 1. Note that this solution is equatorially symmetric.

We see from (3.2) that the surface $b = ak \tanh k$ corresponds to a null streamfunction field. This surface separates thermally (TH) and salinity-(SA) driven symmetric circulations in the space of the control parameters. In fact this separation extends to arbitrary large values of a and b, as can easily be seen in the equations, since (3.1) and (3.2) also describe a stationary solution when only Ψ is small (i.e., in the vicinity of $b = ak \tanh k$). The particular example in Figure 1 is of type SA.

3.2 Three competing stable equilibria

When the forcing is increased (i.e., a and b are no longer small), the nonlinear terms usually cannot be neglected, and multiple steady states can occur. With a numerical solution of (2.3), we are able to reach three different types of stable steady states for equal values of the control parameters. Since the phenomenon is present in the

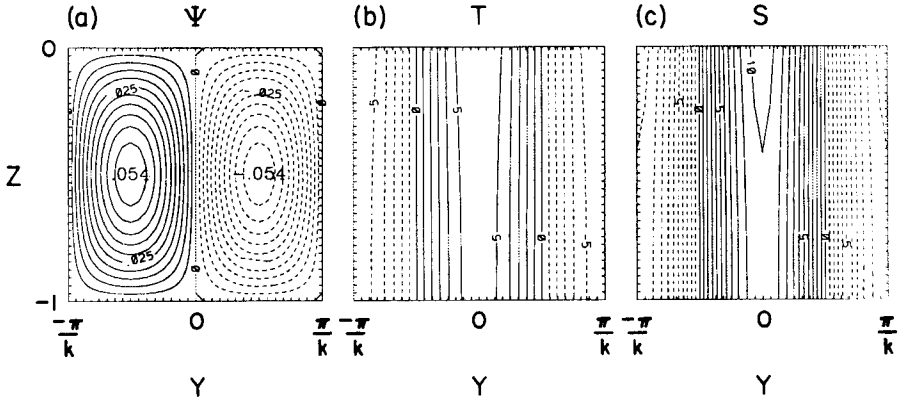

Figure 1 Diffusive regime for the fluid model at small surface forcing ($a = 6$, $b = 1.6$), given by the analytical expressions (3.1) and (3.2). We have chosen $k = 0.4$ to match Figure 2. (a): Streamfunction $\Psi(y, z)$, with a contour interval of 0.005, (b): Temperature $T(y, z)$ and (c): Salinity $S(y, z)$, with contour intervals of 1 for the scalar fields.

limit $\sigma = \infty$, we deduce that the temperature and salinity advection terms are the pertinent nonlinear terms responsible for this effect.

Figure 2 shows a typical example of multiple stable steady states. A thermally driven symmetric equilibrium, denoted by TH, consists of formation of deep water at the poles and upwelling at the equator. Notice the strong circulation, the occurrence of a thermocline in the equatorial half of the domain, and the weak salinity gradients. A salinity-driven symmetric circulation, SA, has circulation in the opposite sense. The gradients of its temperature and salinity fields are more horizontal than in TH, and its circulation is very much weaker. The third type of steady state is a pole-to-pole circulation, denoted by PP, which consists of downwelling at one pole and upwelling at the other pole. There are in fact two such steady states PP because of the equatorial

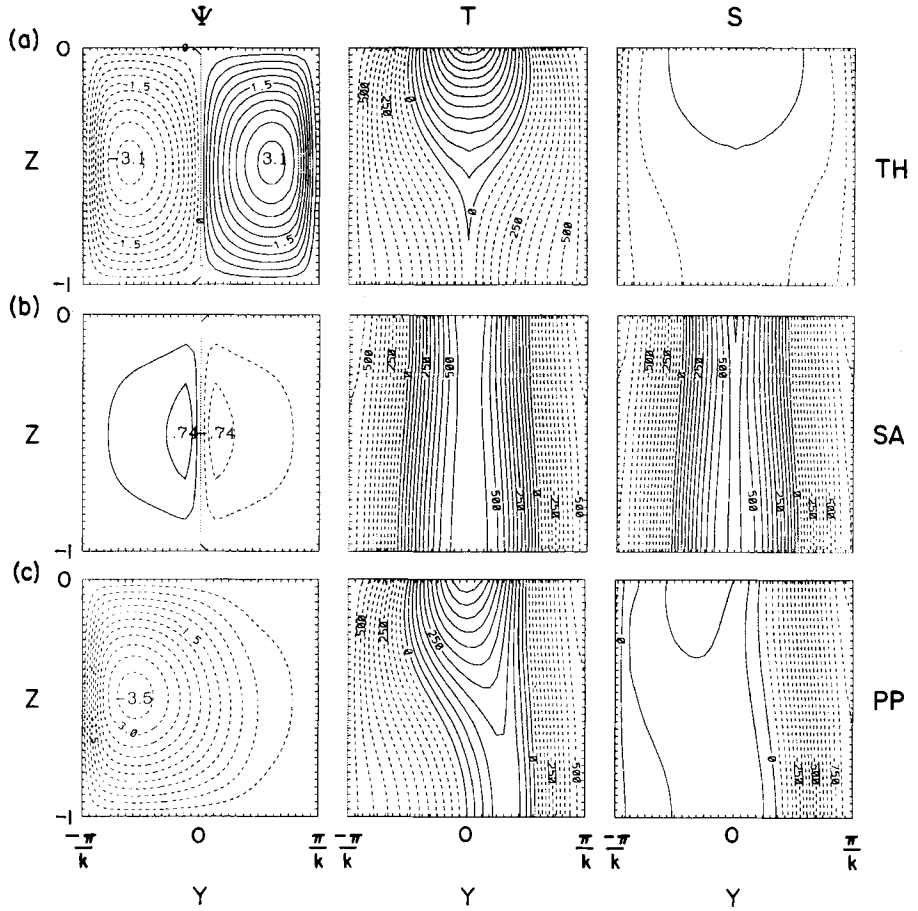

Figure 2 Three stable equilibria for the fluid model. The parameters are $a = 600$, $b = 160$, $k = 0.4$, $\tau = 1$, $\sigma = \infty$. Contour intervals are 0.3 for the stream function Ψ and 50 for the scalar fields T and S. (a): Thermally driven circulation TH. (b): Salinity-driven circulation SA. (c): Pole-to-pole circulation PP.

symmetry of the problem. By looking at these countour plots, one can qualitatively interpret this PP steady state as the "superposition" of half a TH symmetric circulation (southern hemisphere for Figure 2) and half a SA symmetric circulation (northern hemisphere). For a rough correspondence with the present state of the Atlantic Ocean, one should reverse north and south in this PP solution.

3.3 The symmetric and asymmetric catastrophe surfaces

When we follow an experimental path in the control space (e.g., varying b and fixing all the other parameters as indicated in Figure 3c), we obtain a bifurcation diagram for the equilibria, which looks schematically like Figure 3ab. This diagram is drawn from a collection of numerical experiments in which only the stable states could be reached (with or without imposing equatorially symmetry). We infer the existence and qualitative structure of the unstable states, which connect together the known stable states, by likehood arguments. We have not built a non-evolutionary, steady-state solver to compute these unstable states.

When equatorial symmetry is imposed, the two stable symmetric equilibria SA and TH destabilize through saddle-node bifurcations for the values f_1 and f_2 of the varied experimental parameter (e.g., b). When asymmetric perturbations are allowed, SA and TH can be destabilized through symmetry-breaking pitchfork bifurcations at g_1 and g_2. These bifurcations are subcritical (i.e., both of the bifurcated equilibria are unstable). The bifurcated asymmetric states restabilize through saddle-node bifurcations, at e_1 and e_2 to the observed stable pole-to-pole equilibria PP. We have observed that these symmetry-breaking bifurcations at g_1 and g_2 are slow instabilities compared to the symmetric saddle-node bifurcations at f_1 and f_2: the critical real eigenvalue, representing the growth rate of modal perturbations, crosses the zero value more slowly when varying the control parameters. It is thus time consuming to determine exactly the position of these pitchfork bifurcations by direct integration of the model, and, except for few cases, we have achieved only a crude estimation. But we have observed cases (e.g., Figure 2) for which the relative positions of g_1 and g_2 are as shown in Figure 3, allowing the competition between two stable symmetric states and the asymmetric stable states, hence three stable equilibria for same values of (a, b).

If we now consider several experimental paths (e.g., several values of a), we can draw the loci of the saddle-node bifurcations, which are codimension 1 surfaces in the control space. They are four curves f_1, f_2, e_1 and e_2 in a (a, b) projection of the control space, as drawn schematically on Figure 3c. In the language of catastrophe theory, they correspond to folds, which are codimension 1 catastrophes. Also shown are the pitchfork bifurcation curves g_1 and g_2, which in fact must belong to the same differentiable curve g. Its shape and position with respect to f_1 and f_2 are the simplest configuration consistent with our experiments.

The intersection of two fold catastrophes, a cusp, is a codimension 2 catastrophe. It is assumed in Figure 3c that there are two cusps, $p = f_1 \cap f_2$ for the symmetric equilibria and $q = e_1 \cap e_2$ for the asymmetric equilibria. But the curves e_1 and e_2 could end by intersecting tangentially the curve g: the domain of existence of the

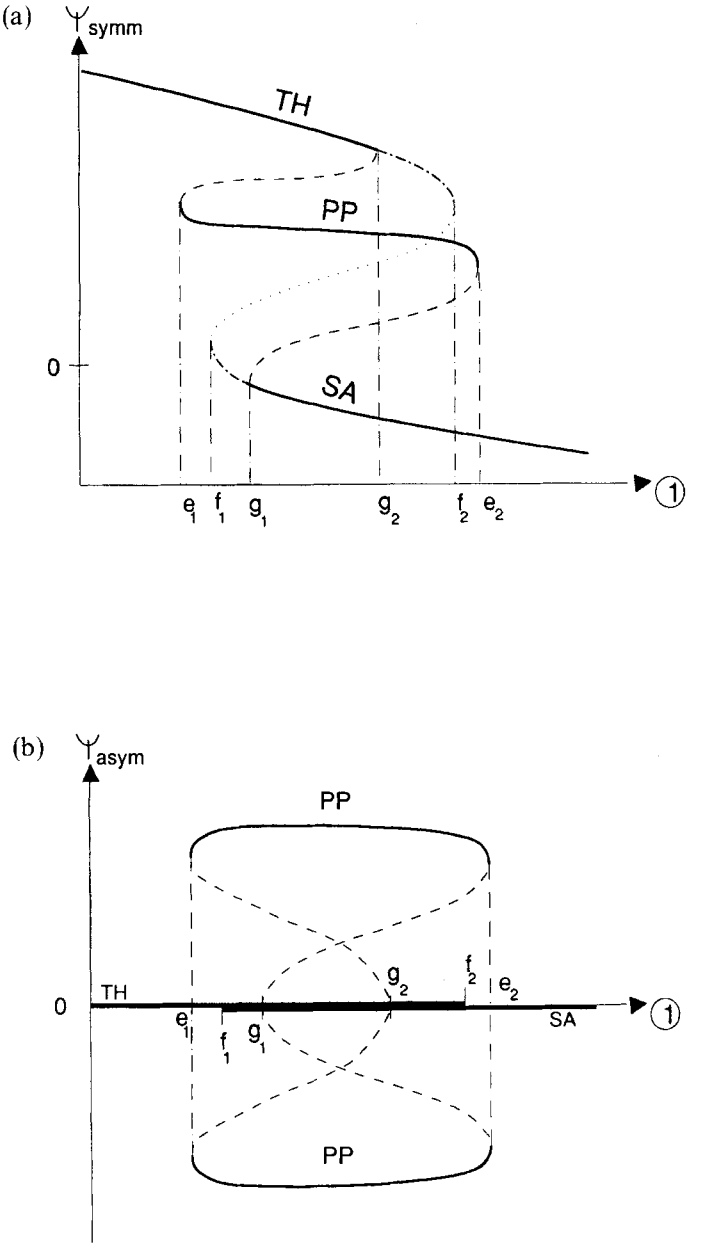

Figure 3 Schematic representation of the catastrophe structure for the steady states of the fluid model. (a, b): Schematic bifurcation diagrams for a generic experimental path in the control space. ———: stable steady state. – – – – –: unstable steady state. –·–·–·–: unstable to asymmetric purturbation only. ·····: unstable to both asymmetric and symmetric perturbations. Ψ_{symm}(a) and Ψ_{asym}(b) are respectively some measures of the symmetric and asymmetric projections of the streamfunction.

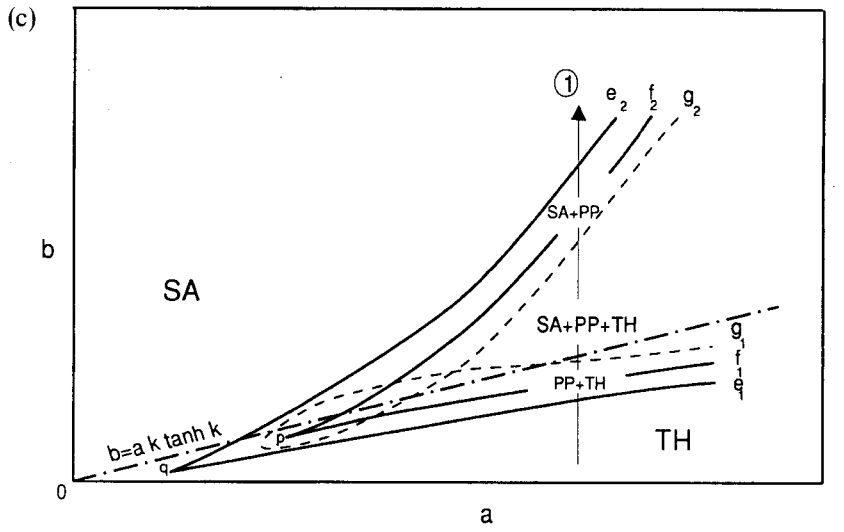

Figure 3 (cont.) (c): Schematic locations of the symmetric (f_1, f_2) and asymmetric (e_1, e_2) folds and cusps in the (a, b) plane.

pole-to-pole circulation would then look like a "smoothed cusp". Our numerical exploration of the control parameters does not have fine enough resolution to decide whether the asymmetric cusps are "smooth" or "sharp".

We show in Figure 4 an experimental determination of the symmetric and asymmetric cusps in the (a, b) plane, for two values of k. We note that as k decreases, both cusp points p and q come increasingly close to the zero-circulation line, and they also appear to converge towards $(a, b) = 0$. Also as $k \to 0$, the lower branches of the symmetric and asymmetric cusps converge toward the zero-circulation line, which itself converges toward the $b = 0$ line. In this limit ($k \to 0$), the symmetric and asymmetric cusps join together. Encouraged by the previously stated, qualitative interpretation of Figure 2, we can conjecture a "superposition principle" to be strictly valid at small k: a stable pole-to-pole circulation PP is found whenever a symmetric thermally driven stable state TH and a symmetric salinity-driven stable state SA can be found for the same value of the control parameters.

We have also considered the restoring and mixed boundary conditions model (see Section 1) where, instead of imposing the present cosine surface salinity flux, we first impose a cosine surface salinity, compute the associated steady state and its surface salinity flux, and then impose the latter for the mixed boundary condition. By this procedure we have also found multiple equilibria, but we do not report here the catastrophe structure of this model. Nevertheless, the two models coincide when $\tau = 1$ and $b = ak \tanh k$, i.e., when the thermal and saline effects can balance exactly. Indeed, for these values of the parameters, the diagnosed salinity flux is a cosine function and the temperature and salinity fields are given by (3.1) with $\Psi = 0$. The three stable states TH, SA and PP which are found at these values of the parameters are thus common to the two models.

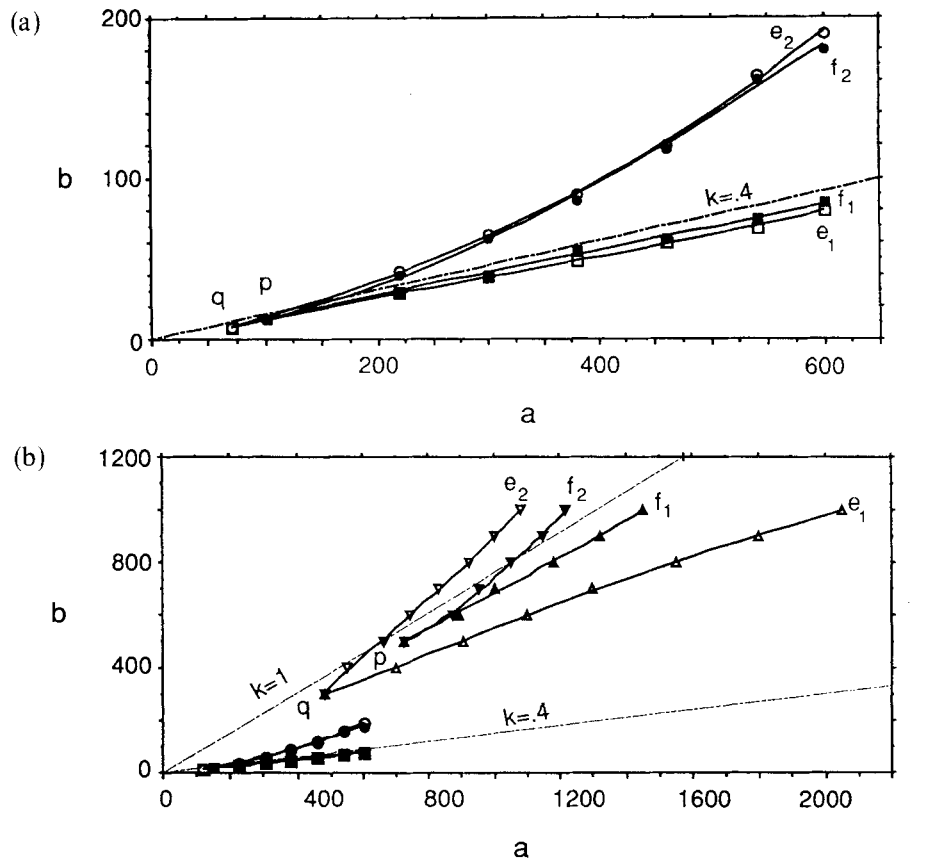

Figure 4 Numerically determined cusps for the fluid model, in the (a, b) plane. Solid symbols indicate the symmetric cusps, and open symbols indicate the asymmetric cusps. Dashed lines are zero-circulation surfaces. (a): Symmetric (f_1, f_2) and asymmetric (e_1, e_2) cusps for $k = 0.4$. (b): Comparison of the cusp structures between $k = 1$ and $k = 0.4$.

4. THE SYMMETRIC STATES AND THE TWO-BOX MODEL

The simplest box model of thermohaline circulation reproduces the qualitative catastrophe structure of the fluid model. We focus here on the symmetric solutions which can be described with only two boxes.

4.1 The two-box model

We use the original two-box model studied by Stommel (1961), in which two well mixed reservoirs communicate through capillary pipes (i.e., with flow rate proportional to the forcing potential, here the density difference); see Figure 5. The model assumes a linear relaxation of the temperature and salinity to external values, in conjunction with nonlinear exchange between the boxes through the capillary pipes.

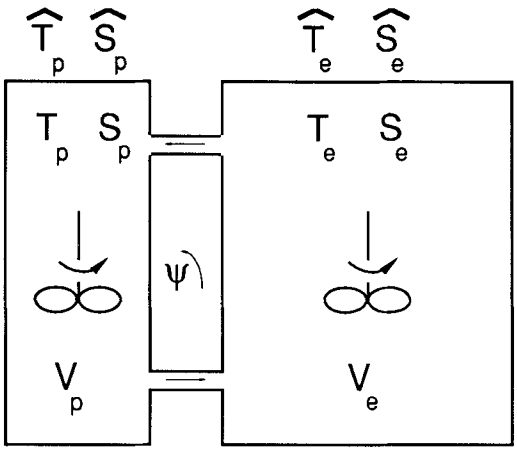

Figure 5 Configuration of the two-box model.

Let V_p, T_p and S_p be the volume, temperature and salinity of the polar box; we use the index e for their counterparts in the equatorial box. We suppose that the circulation Ψ in the capillary tubes, taken positive when it is thermally driven, is proportional to the temperature and salinity differences:

$$\Psi = D[\gamma_T(T_e - T_p) - \gamma_S(S_e - S_p)]. \tag{4.1}$$

The model equations are

$$\begin{aligned} V_p\dot{T}_p &= C_{Tp}(\hat{T}_p - T_p) + |\Psi|(T_e - T_p), \\ V_e\dot{T}_e &= C_{Te}(\hat{T}_e - T_e) + |\Psi|(T_p - T_e), \end{aligned} \tag{4.2}$$

$$\begin{aligned} V_p\dot{S}_p &= C_{Sp}(\hat{S}_p - S_p) + |\Psi|(S_e - S_p), \\ V_e\dot{S}_e &= C_{Se}(\hat{S}_e - S_e) + |\Psi|(S_p - S_e). \end{aligned} \tag{4.3}$$

The relaxation times to the forcing are given by the inverse of the coefficients C_{Ti} for the temperature and C_{Si} for the salinity. We will assume that these coefficients are proportional to the volumes, which implies that both $R_T = C_{Ti}/V_i$ and $R_S = C_{Si}/V_i$ are each the same for the two boxes. We will call $\xi = R_S/R_T = C_{Si}/C_{Ti}$.

The advection terms in (4.2) and (4.3) do not depend on the actual sign of the circulation. Their expression may be explained by a simple mixing argument. In the time δt, an amount of fluid of volume $\delta V = |\Psi|\,\delta t$, at the temperature T_e, is injected in the box p. It mixes with the volume $V_p - \delta V$ to reach the new temperature $\mu T_e + (1 - \mu)T_p$ where $\mu = \delta V/V_p$. The same argument applies for the injection of fluid into the box e, to conserve mass, so that the net effect is the same no matter what the sign of Ψ.

4.2 Two-box model catastrophe structure

In the two-box model the sums of the temperatures and the salinities relax linearly to equilibrium values (e.g., $\hat{T}_e + \hat{T}_p$), since the advection terms of (4.2) and (4.3) cancel when added. The more interesting variables are the differences $\Theta = T_e - T_p$ and $\Sigma = S_e - S_p$. We choose the following units for non-dimensionalization here: $1/R_T$ for time, $(1/V_p + 1/V_e)^{-1} R_T / D\gamma_T$ for temperature, $(1/V_p + 1/V_e)^{-1} R_T / D\gamma_S$ for salinity, and $(1/V_p + 1/V_e)^{-1} R_T$ for streamfunction. The dimensionless model is thus

$$\begin{aligned} \dot{\Theta} &= \alpha - \Theta - |\Psi|\Theta, \\ \dot{\Sigma} &= \beta - \xi\Sigma - |\Psi|\Sigma, \\ \Psi &= \Theta - \Sigma, \end{aligned} \tag{4.4}$$

where $\xi = R_S/R_T$, $\alpha = \hat{T}_e - \hat{T}_p$ and $\beta = \xi(\hat{S}_e - \hat{S}_p)$ are the three dimensionless control parameters. We see that V_e/V_p is not a relevant control parameter after non-dimensionalization.

The equilibria of this simple model are given by $\Theta = \alpha/(1 + |\Psi|)$ and $\Sigma = \beta/(\xi + |\Psi|)$ and the solution of the implicit equation for Ψ:

$$\Psi(1 + |\Psi|)(\xi + |\Psi|) = \alpha(\xi + |\Psi|) - \beta(1 + |\Psi|). \tag{4.5}$$

For fixed ξ the zero-circulation line in the plane (α, β) is $\beta = \xi\alpha$. It is possible to determine analytically the domain for which multiple states exist (Figure 6a). For positive α (negative α is obtained by a symmetry argument), this domain has a cusp-like structure, starting from the point $p = [\xi/(1 - \xi), \xi^2/(1 - \xi)]$ on the zero-circulation line, and lying between the zero-circulation line, which can be seen as f_1, and a curve f_2 whose analytical expression is given in Appendix B. We note that the contact (i.e., the rate of convergence between the two curves) at p between the fold f_2 and the zero-circulation line has a quadratic dependence on distance away from the intersection point, instead of the usual exponent of $\frac{2}{3}$ for a cusp as predicted by the catastrophe theory for differential systems (Arnold, 1984). This discrepancy is due to the non-differentiability of the box model coming from the $|\Psi|$ term. However it is striking to notice that the cusp of the fluid model (Figure 4) looks like the one of the two-box model on the scale on which we have resolved the fold lines (i.e., on which the figure is drawn).

When $\xi \to 0$, the zero-circulation line converges to the line $\beta = 0$, the point p converges to $(0, 0)$, and the fold f_2 converges to the parabola $\beta = \alpha|\alpha|/4$, as can be seen from its analytic expression (Appendix B) or from the asymptotic expansion of Sections 5.2 and 5.3 (Figure 9b).

4.3 Connection with the fluid model

It is heuristically plausible that the pipe circulation of the two-box model mimics the thermohaline circulation of the 2D Boussinesq model. However, the connection between the forcings of these models needs some explanation. The departure of the

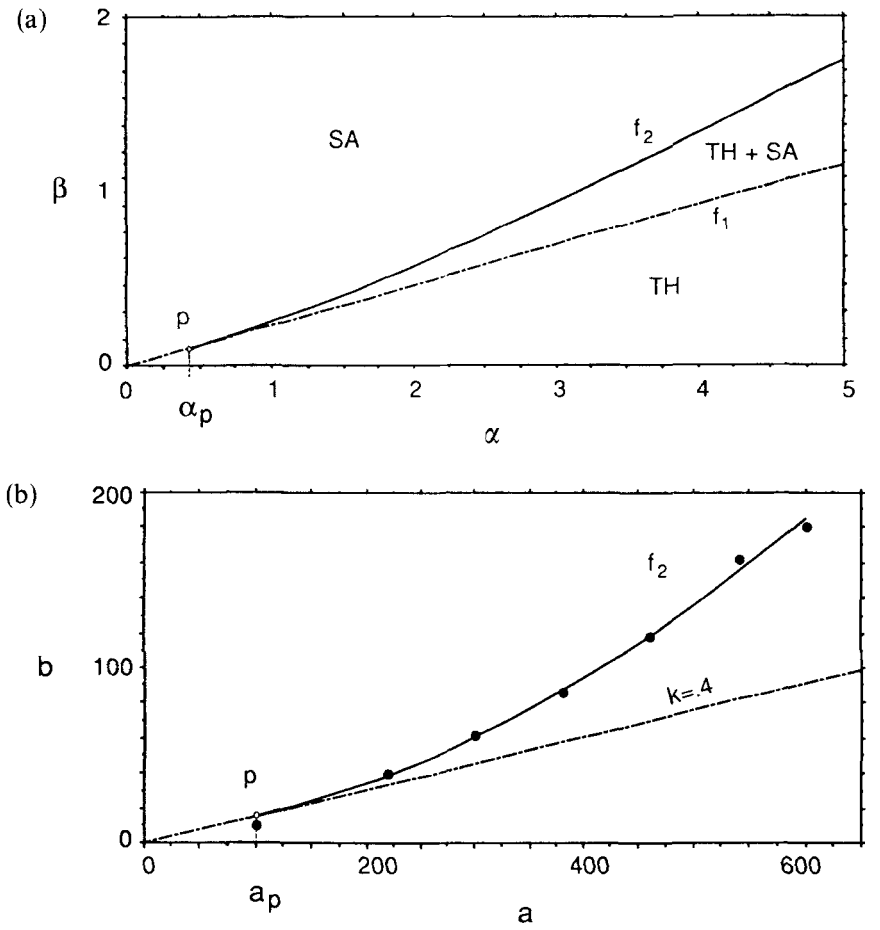

Figure 6 Two-box model catastrophe structure from its analytical expression. (a): For $\xi = 0.3$ in the (α, β) plane. (b): For $\xi = 0.002$ in a (a, b) plane obtained by the scaling $a = 0.5\ 10^4\alpha$, $b = 3.8\ 10^6\beta$, to fit the experimental points of the $k = 0.4$ fluid model as described in the text.

quantity ξ from 1 (in particular, $\xi < 1$) is intended to mimic the difference in the nature of the temperature and salinity forcing for the fluid model: it is plausible that a scalar field relaxes faster when its value is imposed at a boundary (here fixed temperature) than when its flux is imposed (here fixed salinity flux). Thus ξ must be taken to be a small parameter to match the mixed boundary conditions of the fluid model, and this explains our choice of β as a control parameter which mimics the intensity of the salinity forcing independently from the choice of ξ. There is no difficulty in interpreting α as the intensity of the temperature forcing.

In order to find a better connection with the Boussinesq model, we can enrich the two-box model by introducing a horizontal diffusivity κ between the two boxes. This consists in changing $|\Psi|$ into $|\Psi| + \kappa$ in (4.4). However this new parameter can be

removed by rescaling the variables and the time in the following manner: $\Theta' = \Theta/(1+\kappa)$, $\Sigma' = \Sigma/(1+\kappa)$, and $t' = t(1+\kappa)$. The original form of the equation (with $\kappa = 0$) is recovered by renaming the following quantities: $\alpha' = \alpha/(1+\kappa)^2$, $\beta' = \beta/(1+\kappa)^2$ and $\xi' = (\kappa+\xi)/(1+\kappa)$. Thus, the slope ξ' of the zero circulation line in a (α', β') plane is controlled by two physical effects: the ratio ξ of relaxation times to forcing and the ratio κ between horizontal diffusion and advection magnitude.

With this enrichment, we can compare the catastrophe structures of the symmetric steady states of the fluid model and the two-box model. The cusp point p and the slopes of the zero-circulation lines, $k \tanh k$ for the fluid model and ξ for the box model—or $\xi' = (\kappa+\xi)/(1+\kappa)$ if a horizontal diffusivity κ is included—provide a first qualitative correspondence between the control parameters of the two systems. For small values of these slopes, the curve f_1 of the fluid model converges toward the zero-circulation line (the two lines are the same in the two-box model for all values of ξ), the cusp point p of both models converge toward the origin, and the curves f_2 of both models converge towards a parabola passing through the origin—i.e., for the two-box model the asymptotic curve is $\beta = \alpha|\alpha|/4$ (Figure 9b), and for the fluid model f_2 has positive curvature, approximately parabolic, for small k (Figure 4a).

For a given small value of k, a quantitative correspondence between the control parameters of the fluid model (a, b, k) and those of the two-box model (α, β, ξ) can be made by setting $\alpha = A\alpha$, $b = B\beta$ and $k \tanh k = C\xi$. We can require that the abscissa a_p of the fluid model cusp point p corresponds to the abscissa α_p of the two-box model p point, which leads to the first condition $a_p = A\xi/(1-\xi)$. It is natural to impose that the zero-circulaton lines correspond, which simply implies $C = B/A$. (We cannot match both the abscissa and ordinate of p and the zero-circulation line, since these are distinct in the fluid model at finite k and coincident in the two-box model for all ξ.) The third and last condition in order to complete the correspondence can be achieved with a nonlinear regression of the fluid model fold f_2 by the analytically known, two-box model curve f_2. We have performed this correspondence for $k = 0.4$, by a nonlinear regression with seven experimental points on the fluid model fold f_2, and found a best fit for $\xi = 0.002$, which leads to $A = 5.0 \times 10^4$, $B = 3.8 \times 10^6$ and $C = 76$. Figure 6b shows how well the rescaled two-box model curve f_2 fits the experimental points from the fluid model in the plane (a, b). The low value of ξ which comes out of this correspondence procedure reinforces the physical interpretation of this parameter given above.

5. ASYMMETRIC STATES AND THE THREE-BOX MODEL

We generalize the two-box model to a three-box model, in order to investigate the catastrophe structure when asymmetric solutions are permitted. This model yields the "superposition principle" exactly in the limit $\xi \to 0$.

5.1 The three box-model

We now consider two polar boxes (index s for the southern box and n for the northern

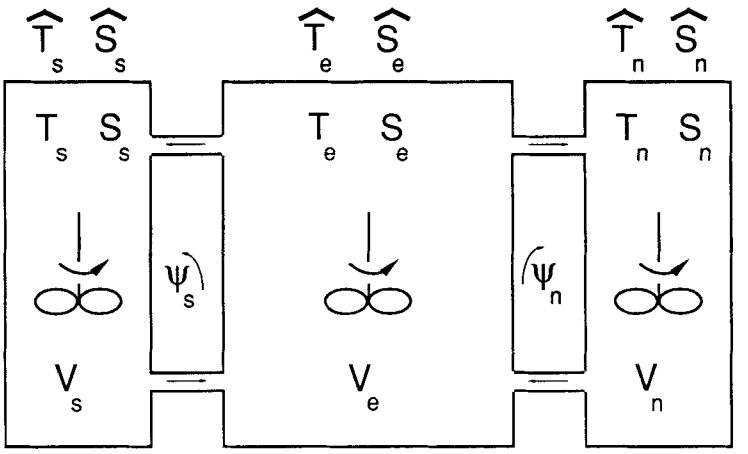

Figure 7 Configuration of the three-box model.

box) surrounding the equatorial box (Figure 7). We call $\Theta_s = T_e - T_s$, $\Theta_n = T_e - T_n$, $\Sigma_s = S_e - S_s$ and $\Sigma_n = S_e - S_n$ the temperature and salinity differences (again the mean temperature and salinity simply relax to the averages of the forcing values). The circulations in the pipes Ψ_s and Ψ_n are governed by the same law (4.1) as for the two-box model. We assume that the two polar boxes have the same volumes and are forced with the same intensities. This equatorially symmetric forcing involves only two non-dimensional parameters, viz., $\alpha = \hat{T}_e - \hat{T}_s = \hat{T}_e - \hat{T}_n$ and $\beta = \xi(\hat{S}_e - \hat{S}_s) = \xi(\hat{S}_e - \hat{S}_n)$.

For simplicity we choose the equatorial box volume to be twice the polar box volumes, but our analysis can be done the same way, with the same conclusions, for any other choice. After a suitable non-dimensionalization we obtain the following three-box model:

$$\begin{aligned}
\dot{\Theta}_s &= \alpha - \Theta_s - \tfrac{3}{4}|\Psi_s|\Theta_s - \tfrac{1}{4}|\Psi_n|\Theta_n \quad , \\
\dot{\Theta}_n &= \alpha - \Theta_n - \tfrac{3}{4}|\Psi_n|\Theta_n - \tfrac{1}{4}|\Psi_s|\Theta_s, \\
\dot{\Sigma}_s &= \beta - \xi\Sigma_s - \tfrac{3}{4}|\Psi_s|\Sigma_s - \tfrac{1}{4}|\Psi_n|\Sigma_n, \\
\dot{\Sigma}_n &= \beta - \xi\Sigma_n - \tfrac{3}{4}|\Psi_n|\Sigma_n - \tfrac{1}{4}|\Psi_s|\Sigma_s, \\
\Psi_s &= \Theta_s - \Sigma_s \quad \text{and} \quad \Psi_n = \Theta_n - \Sigma_n.
\end{aligned} \tag{5.1}$$

The symmetric solutions of this model are the solutions of the two-box model: $\Theta_n = \Theta_s = \Theta$, $\Sigma_n = \Sigma_s = \Sigma$, and $\Psi_n = \Psi_s = \Psi$ from (4.4).

The analytical determination of the asymmetric fold surfaces of the three-box model seems cumbersome. Figure 8 shows a numerical determination of the catastrophe

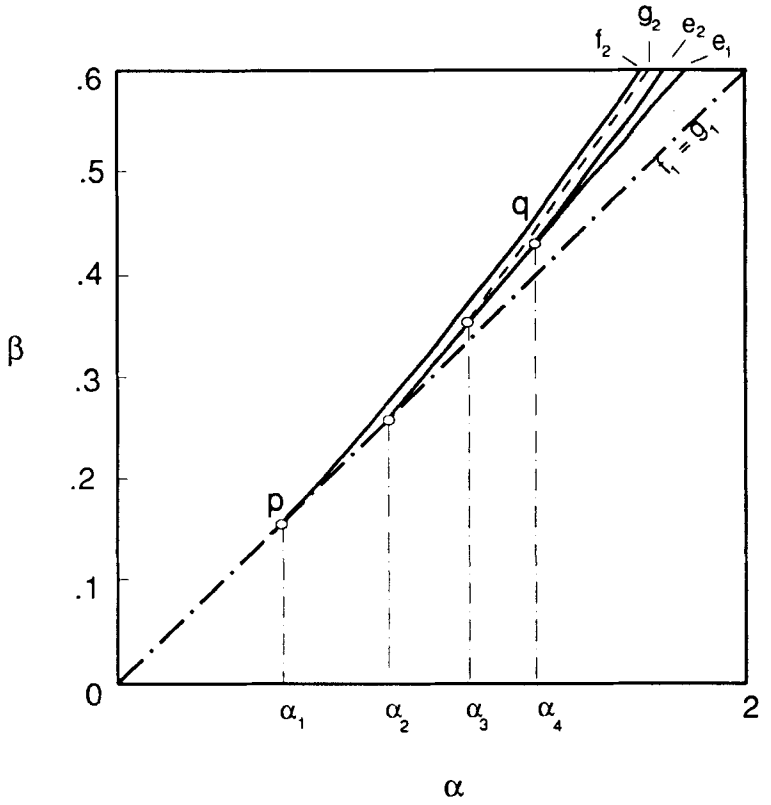

Figure 8 Three-box model catastrophe structure in the (α, β) plane for $\xi = 0.3$. The curves f_1 and f_2 are those of the two-box model (see Figure 6). The curves e_2 and g_2 have been determined numerically; the curve e_1 can be determined analytically. The zero-circulation line, f_1, and g_1 are the same. All these curves have been stopped at intersection points, which abscissas are denoted by the α_i's.

structure of the three-box model, for a typical value of ξ. The curves f_1 (also the zero-circulation line) and f_2 are the same as in the two-box model, and they delimit the domain of competing symmetric equilibria. The curves e_1 and e_2 delimit the domain of existence of stable asymmetric equilibria. The curve e_1 can be determined analytically by requiring that $\Psi_s = 0$ (or equivalently $\Psi_n = 0$) in (5.1). The fold e_2, corresponding to a saddle-node bifurcation of pole-to-pole equilibria, and the curve g_2, corresponding to a symmetry breaking, pitchfork bifurcation of the unstable symmetric state, have been determined numerically by the "software for continuation and bifurcation problems" called AUTO (Doedel, 1981). Finally, we set $g_1 = f_1$ to indicate that an asymmetric equilibrium originates at the zero-circulation line, as can be shown by an asymptotic expansion, or numerically with AUTO. Time-dependent solutions of the dynamical system (5.1) show that the basins of attraction of the pole-to-pole equilibria are quite small in this three-box model.

As for the two-box model, the physical interpretation of the model can be enriched

by introducing a horizontal diffusivity κ, which can be absorbed into the three control parameters by a suitable rescaling. The same connection with the fluid model as the one developed in Section 4 can be made, since the catastrophe structure of the symmetric states is exactly that of the two-box model. But the general catastrophe structure is not completely satisfactory for the three-box model, most strikingly because the positions of e_1 and e_2 do not match those of the fluid model and lie within the folds f_1 and f_2 rather than outside (Figure 4). We will see in Section 6 that this defect can be corrected by considering more complex box models. Nevertheless the three-box model is useful in the limit $\xi \to 0$, since it enables us to understand analytically the "superposition principle" (Section 5.2).

5.2 *Asymptotic expansion for small* ξ

We will concentrate on an asymptotic path in the control space, where $\xi \sim \varepsilon^2$, $\alpha \sim \varepsilon$, and $\beta \sim \varepsilon^2$. We can show that the dependent variables obey the scalings $\Theta_i \sim \varepsilon$, $\Sigma_i \sim \varepsilon$ and $\Psi_i \sim \varepsilon$ for $i \in \{s, n\}$. At leading order in ε the equations are simplified into

$$\begin{aligned} \dot{\Theta}_s &= \alpha - \Theta_s, \\ \dot{\Theta}_n &= \alpha - \Theta_n, \\ \dot{\Sigma}_s &= \beta - \tfrac{3}{4}|\Psi_s|\Sigma_s - \tfrac{1}{4}|\Psi_n|\Sigma_n, \\ \dot{\Sigma}_n &= \beta - \tfrac{3}{4}|\Psi_n|\Sigma_n - \tfrac{1}{4}|\Psi_s|\Sigma_s, \\ \Psi_s &= \Theta_s - \Sigma_s \quad \text{and} \quad \Psi_n = \Theta_n - \Sigma_n. \end{aligned} \tag{5.2}$$

The choice $\xi = \varepsilon^2$ is only governed by aesthetic reasons, since we can take any smaller value to obtain (5.2). We remark that the temperature dynamics decouple from the equations, and that the temperatures converge exponentially to the values $\Theta_n = \Theta_s = \alpha$. This simplified model can be seen as a "fixed temperature and salinity flux" three-box model, and it has often been presented as such in the literature (Welander, 1988; Marotzke, 1989; Martozke, 1990), without reference to this asymptotic path.

5.3 *The superposition principle*

The stationary solutions of the simplified model (5.2) are given by the system

$$\begin{aligned} 0 &= \beta - \tfrac{3}{4}|\Psi_s|\Sigma_s - \tfrac{1}{4}|\Psi_n|\Sigma_n \\ 0 &= \beta - \tfrac{3}{4}|\Psi_n|\Sigma_n - \tfrac{1}{4}|\Psi_s|\Sigma_s, \\ \Psi_s &= \alpha - \Sigma_s \quad \text{and} \quad \Psi_n = \alpha - \Sigma_n, \end{aligned} \tag{5.3}$$

which factorizes into two uncoupled equations, $\beta = |\Psi_s|\Sigma_s$ and $\beta = |\Psi_n|\Sigma_n$. Thus,

they can be "superposed", with separate solutions in each half of the domain if multiple equilibria exist. These half-domain solutions are given by the implicit equation $Y = |X|(1 - X)$, where $Y = \beta/(\alpha|\alpha|)$ and $X = \Psi/\alpha$. For $0 < Y < 1/4$, there are three solutions (Figure 9a). Two of them correspond to stable equilibria: $\Psi = \alpha X^{(3)}$ which is thermally driven (TH), and $\Psi = \alpha X^{(1)}$ which is salinity driven (SA). In between, an unstable equilibrium $\Psi = \alpha X^{(2)}$ connects them through a saddle-node bifurcation at $Y = 1/4$ and another bifurcation at $Y = 0$, where the solution curve is non-differentiable (Figure 9a).

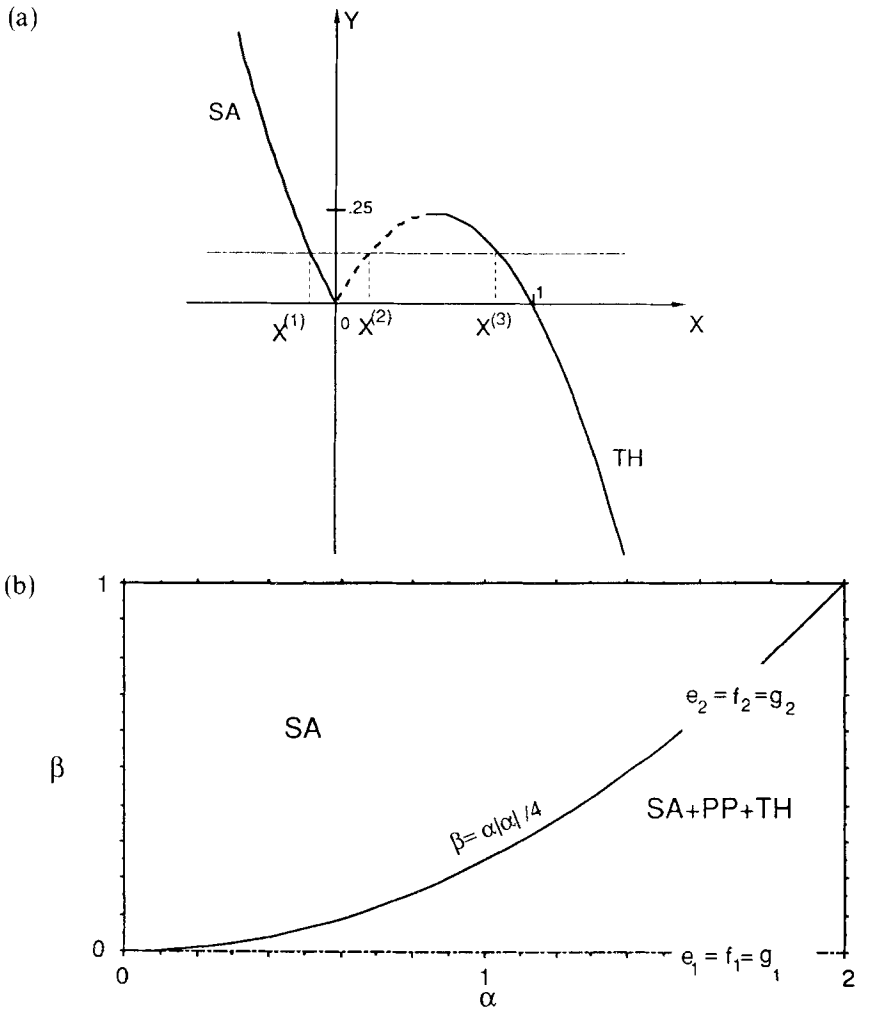

Figure 9 Asymptotic three-box model. (a): Graphical solution of the equation $Y = |X|(1 - X)$. ——: solutions $X^{(1)}$ (SA) and $X^{(3)}$ (TH) which correspond to stable half-domain equilibria. · · · · · · : solution $X^{(2)}$ corresponding to the unstable half-domain equilibrium. (b): Catastrophe structure in the control space (α, β).

Thus, if we do not impose equatorial symmetry, there are nine equilibria $(\Psi_s, \Psi_n) = (\alpha X^{(m_s)}, \alpha X^{(m_n)})$ where $(m_s, m_n) \in \{1, 2, 3\}^2$. We can view these equilibria as the exact "superposition" of two half-domain equilibria. This "superposition principle" also applies when dealing with the stability of these equilbria (see Appendix C). We thus conclude that the four equilibria TH = $(\alpha X^{(3)}, \alpha X^{(3)})$, SA = $(\alpha X^{(1)}, \alpha X^{(1)})$ and PP = $(\alpha X^{(3)}, \alpha X^{(1)})$ or $(\alpha X^{(1)}, \alpha X^{(3)})$ are stable. The five other equilibria, containing $X^{(2)}$ in at least one half-domain, are unstable.

In a plane (α, β), the symmetric and asymmetric cusps are identical, and they are the curve $\beta = \alpha|\alpha|/4$ and the zero-circulation line $\beta = 0$ (Figure 9b). There is a thus an exact "superposition principle" for this simplified (or asymptotic) three-box model: whenever a stable state TH and a stable state SA coexist for a particular value of the control parameter, a stable state PP, with TH in one half-domain and SA in the other, also coexists.

5.4 Asymptotic path for the fluid model

One would like to repeat this asymptotic procedure for the 2D Boussinesq model, to demonstrate rigorously a similar "superposition principle". The numerical experiments suggest that this happens in the limit $k \to 0$. We have tried to build such an asymptotic expansion by choosing the scaling $a \sim \varepsilon$ and assuming that $\Psi \sim \varepsilon$ and $T \sim \varepsilon$. The proper scaling for b and S is more tricky, since it involves boundary layers in y at the poles, as we have observed in numerical solutions for small k. These boundary layers come from the fact that the operator ∇^2 with flux boundary conditions becomes singular in the limit $k \to 0$. Moreover, these boundary layers exhibit dynamical instabilities (e.g., a Hopf bifurcation characterized by the oscillation of a narrow but deep counter-rotating cell at each pole). We have not gone further in this direction.

6. A HIERARCHY OF BOX MODELS

We define a family of different box-model by building a hierarchy, in which the simpler models of Sections 4 and 5 are deduced from more complex ones in some asymptotic limit. We have shown that the three-box model was able to reproduce some qualitative features of the catastrophe structure of the fluid model, but that some others were unsatisfactory (see Section 5.1). We study here slightly more complex box models in order to find the simplest model which reproduces this catastrophe structure in a satisfactory manner. For conceptual simplicity we prefer to investigate a few ocean "boxes" rather than make a low-order Galerkin projection of the fluid equations on suitably differentiable basis functions. Our goal is physical and mathematical understanding of the fluid dynamics by analogy.

6.1 The 3 × 2 box model

The configuration of a 3 × 2 box model is shown in Figure 10. Neighboring boxes are connected by only one pipe. This box and pipe configuration is motivated by

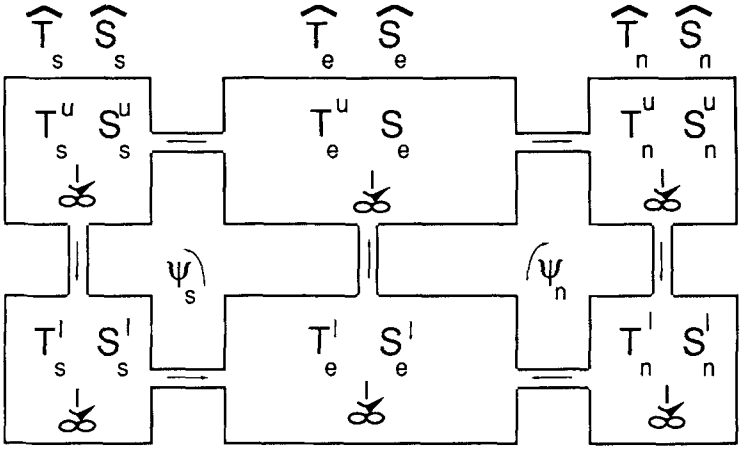

Figure 10 Configuration of the 3 × 2 box model.

the form of the stationary circulations obtained in the fluid model (e.g., Figures 1–2). By defining separate upper and lower boxes, the sign of the circulation is now relevant, and to distinguish signs we define for $i \in \{s, n\}$ the non-negative quantities

$$\Psi_i^+ = \tfrac{1}{2}(|\Psi_i| + \Psi_i),$$
$$\Psi_i^- = \tfrac{1}{2}(|\Psi_i| - \Psi_i). \tag{6.1}$$

As for the two-box and three-box models, we suppose that the pipes are thin, so that a capillary law of the type (4.1) exists, and parcels of fluid entering a reservoir are immediately mixed. We also suppose that there are both horizontal and vertical diffusion through the pipes, whose strengths compared to the advection are measured by the dimensionless parameters κ_h and κ_v, respectively. We call α and β, as usual, the intensities of the temperature and salinity forcing. The dimensionless parameter ξ measures the ratio of the characteristic relaxation times to the forcing of salinity and the temperature; again this parameter must be thought of as a small number in order to mimic the mixed boundary conditions of the fluid model. We call δ the ratio between the lower and the upper box volumes, and we assume, for simplicity, that the equatorial boxes have twice the volume of the polar ones.

As for the two-box and three-box models, we choose a time unit based on the temperature relaxation time to the forcing value and temperature and salinity units based on box volumes, the expansion coefficients γ_T and $-\gamma_S$, and the coefficient D entering in the capillary law. With these choices, the dimensionless equations of the 3 × 2 box model are given in Appendix D. Because of the presence of quantities (6.1), this model is non-analytic, as were the two- and three-box models.

6.2 Simpler models in the hierarchy

A first simplification can be made by imposing equatorial symmetry. Thus the 3×2 model includes a 2×2 model when we look at the symmetric solutions.

The three-box model of Section 5 can be derived as the limit of the 3×2 model when both $\delta \to 0$ and $\kappa_v \to \infty$. This last limit is important because if we only take the limit $\delta \to 0$, additional advection velocities $Q_i = \Psi_i^+ \Psi_j^- / 2(\Psi_i^+ + \Psi_j^- + 2\kappa_v)$ must be included (here i and j are in $\{s, n\}$ with $j \neq i$) in the three box-model equations, leading to the 3×1 model in the present notation. However, these velocities are zero in the symmetric case (i.e., the two-box model of Section 4, or 2×1) since either $\Psi_s^- = \Psi_n^- = 0$ or $\Psi_s^+ = \Psi_n^+ = 0$ when $\Psi_s = \Psi_n$.

One can include a $1 - 2 - 1$ box model in the hierarchy by combining the upper and lower boxes for the polar regions, motivated by the weak polar stratifications observed in the fluid model solutions—this type of box model also has been studied by Marotzke (1990). Formally this is accomplished by adding the upper and lower equations for the polar variables, which are then set equal to each other. The corresponding symmetric model would be called $2 - 1$. An asymptotic link with the other box models of the hierarchy could be established by defining different values for vertical diffusivity κ_v in the polar and the equatorial boxes and letting the polar diffusivity become large.

6.3 Catastrophe structure of the hierarchy

We now explore numerically the catastrophe structures of the box models of the hierarchy. Figure 11 shows a numerical determination of the symmetric and asymmetric cusps for the 3×2 and $1 - 2 - 1$ models, at typical values of ξ, κ_h, κ_b and δ. To determine these cusps we choose initial conditions on a horizontal line ($\alpha = 0$) in the (α, β) plane and let the system relax to an asymptotic state on each point of the line; this yields equilibrium solutions of type TH. We then take these equilibria as initial conditions for a horizontal line slightly above the first one, and so on. After we reach an upper line which is sufficiently high (α large), we reverse the procedure back to the first line. Because of the solution structure (Figure 3), there is a hysteresis such that the two paths, up and down, do not select the same equilibrium in the region between the cusps folds. This procedure is carried out both with symmetry imposed and without it, and in the later case we add a random asymmetric perturbation to each initial condition in order to catch the asymmetric equilibria.

Figure 11 shows that the catastrophe structures of the 3×2 and $1 - 2 - 1$ models are quite similar. Furthermore, they are both qualitatively similar to that of the fluid model (Figure 4), with the asymmetric fold lines outside the symmetric ones. This is a qualitative improvement over the three-box model structure (Figure 8). We do not attempt a quantitative comparison with the fluid model here: given the many parameters available and the success of the match with the two-box model for the symmetric folds (Figure 6b), we would expect to be able to achieve a very close correspondence.

We have investigated a variety of values for the control parameters for each model

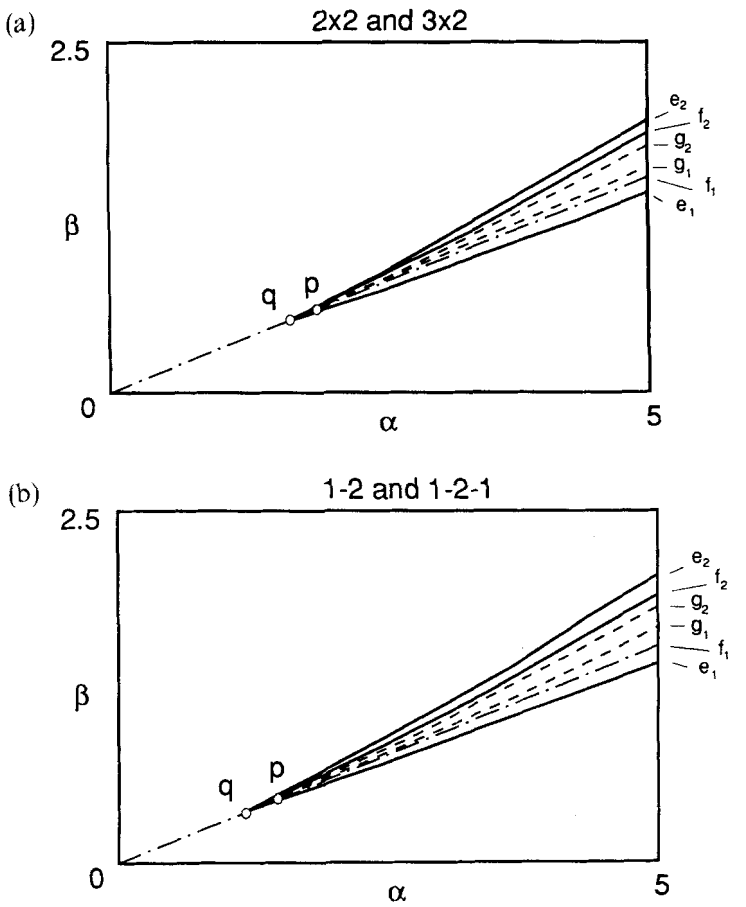

Figure 11 Catastrophe structure of the hierarchy: (a): 3 × 2 model and (b): 1 − 2 − 1 model for $\xi = 0.2$, $\kappa_h = 0.1$, $\kappa_v = 0.1$, and $\delta = 1$ in the (α, β) plane. The curves f_1 and the zero-circulation line are the same. The curves e_1, g_1, g_2, f_2 and e_2 have been determined numerically by the "up and down" procedure (see text).

of the hierarchy. When ξ, κ_h and κ_v are varied, the catastrophe srtructure remains qualitatively the same, with the strongest variations in the zero-circulation line and the folds e_1 and f_1 which follow it. For the 3 × 2 model the slope of the zero-circulation line is

$$\frac{2\delta\kappa_h^2 + (\delta + 1)\kappa_h\kappa_v + \xi(2\delta\kappa_h + \kappa_v)}{2\delta\kappa_h^2 + (\delta + 1)\kappa_h\kappa_v + 2\delta\kappa_h + \kappa_v}. \tag{6.2}$$

We conjecture that the "superposition principle" is true for all these models in an asymptotic path in the control parameter similar to that used for the three-box model (Section 5).

When δ is decreased the asymmetric cusp is no longer visible through the "up and down path" procedure described above, at least in the portion of the (α, β) plane we have explored. This means that the symmetrically stable equilibria SA and TH are no longer destabolized by asymmetric perturbations. But there is still a stable pole-to-pole equilibrium PP, with an associated cusp-catastrophe structure. This behavior when $\delta \to 0$ (i.e., the 3×1 box model) demonstrates why the catastrophe structure of the three-box model does not match well that of the fluid model. At least one lower box, as in the $1 - 2 - 1$ model, is needed to achieve this goal, and a lower box is required to distinguish the effect of the sign of the circulation.

7. CONCLUSIONS

We have explored the control parameters (a, b, k) of a 2D Boussinesq model of oceanic thermohaline convection and found a simple catastrophe structure. We have found two cusps, one for the symmetric and one for the asymmetric equilibria. For small values of k, there appears to be is a "superposition principle" such that, inside the symmetric cusp, the symmetric, stable, thermally (TH) and salinity-(SA) driven circulations linearly combine to form a stable pole-to-pole (PP) circulation. This "superposition principle" explains why, in this limit of small aspect ratio, the cusp for the symmetric solution matches the asymmetric cusp. In this limit one fold of each cusp converges toward the zero-circulation surface, which itself converges to the line $b = 0$. The other folds appear to converge to a common line that resembles a parabola $b \sim a|a|$. The intersection of two cusps is a codimension 3 phenomena, but we have not investigated the exact nature of this higher order catastrophe. Even outside this asymptotic limit, the qualitative validity of this "superposition principle" persists.

Box models are able to reproduce this catastrophe structure. For the two-box model we have determined analytically the symmetric cusp which matches quite closely the symmetric cusp of the fluid model. We have defined a quantitative correspondence with the fluid model which supports the smallness of the parameter β, the ratio of salinity and temperature relaxation times to the forcings, or $\xi' \sim \xi + \kappa$, when a horizontal diffusivity κ is added. For the three-box model we have exhibited the low ξ or ξ' asymptotic limit in which the "superposition principle" can be demonstrated analytically. However the three-box model is not powerful enough to reproduce well the catastrophe structure of the fluid model outside this asymptotic limit: the symmetric cusp is no longer in the interior of the asymmetric one. We have thus embedded these two simple models into a hierarchy of box models, which may appear more physical plausible as an analog of the fluid model. The box models of this hierarchy are related to each other by letting the parameters tend to certain limit values. The introduction of at least one bottom box is sufficient to reproduce qualitatively the full catastrophe structure of the fluid model.

Acknowledgements

We appreciate discussions with Frank Bryan and Jeffrey Weiss during the course of this work. We thank Eusebius Doedel for helping us acquire his software AUTO. This work was supported by the National Science Foundation through its contract with the National Center for Atmospheric Research.

References

Arnold, V. I., *Catastrophe theory*, Springer Verlag (1984).

Arnold, V. I., *Mathematical methods for classical mechanics*, Springer Verlag (1989).

Birchfield, G. E., "A coupled ocean-atmosphere climate model: temperature versus salinity effects on the thermohaline circulation," *Climate Dyn.* **4**, 57–71 (1989).

Broecker, W. S. and Denton, G. H., "The role of ocean-atmosphere reorganizations in glacial cycles," *Geochim. et Cosmochim. Acta* **53**, 2465–2501 (1989).

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

Bryan, F. and Oort, A., "Seasonal variation of the global water balance based on aerological data," *J. Geophys. Res.* **89**, 11717–11730 (1984).

Doedel, E. J., "AUTO: A program for the automatic bifurcation analysis of autonomous systems," *Cong. Num.* **30**, 265–284 (1981).

Huang, R. X., Luyten, J. R. and Stommel, H. M., "Multiple equilibrium states in combined thermal and saline circulation," preprint (1990).

Manabe, S. and Stouffer, R. J., "Two stable equilibria of a Coupled Ocean Atmosphere Model," *J. Climate* **1**, 841–866 (1988).

Marotzke, J., *Instabilities and multiple steady states of the thermohaline circulation, in Oceanic circulation Models: Combining data and dynamics*, Kluwer Academic Publishers, 501–511 (1989).

Marotzke, J., "Instabilities and Multiple Equilibria of the Thermohaline Circulation," Ph.D. Thesis, Christian Albrechts University of Kiel (1990).

Marotzke, J., Welander, P. and Willebrand, J., "Instability and multiple steady states in a meridional-plan model of the thermohaline circulation," *Tellus* **40A**, 162–172 (1988).

Rooth, C., "Hydrology and ocean circulation," *Prog. Oceanog.* **11**, 131–149 (1982).

Stommel, H. M., "Thermohaline convection with two stable regimes of flow," *Tellus* **XIII 2**, 224–230 (1961).

Walin, G., "The thermohaline circulation and the control of ice ages," *Palaeogeogr., Palaeoclimatol, Palaeoecol.* **50**, 323–332 (1985).

Welander, P., "Thermohaline effects in the ocean circulation and related simple models," In: *Large-scale transport processes in oceans and atmosphere* (Eds J. Willebrand and D. T. L. Anderson), D. Reidel, 163–200 (1988).

APPENDICES

A. *Stream function in the diffusive regime*

The stream function vertical profile $\phi(Z)$ of (3.2) is obtained by solving the elliptic problem (2.7) where T and S are given by (3.1), with the free slip boundary conditions $\Psi = \Psi_{zz} = 0$ at $z = 0, -1$. The lateral boundary conditions, either free slip or periodic, are trivially satisfied by the separation of the variables y and z. After cumbersome algebraic manipulation we obtain

$$\phi(Z) = \frac{\cosh k}{8} \{[-P(k)Z + Z^2] \cosh Z + [Q(k) - Z + Z^2 \tanh k] \sinh Z\},$$

with

$$P(k) = \frac{\sinh k^2 - 2 \sinh k(\cosh k - \tanh k \sinh k)k + k^2}{\cosh k \operatorname{siny} k - k},$$

and

$$Q(k) = \frac{k^3 \sinh k(\cosh k - \tanh k \sinh k)}{\cosh k \sinh k - k}.$$

B. Analytical determination of the two-box model catastrophe

The solutions Ψ of (4.5) are roots of two degree 3 polynomials, one polynomial for the case $\Psi \geqslant 0$ and one for $\Psi \leqslant 0$. The fold catastrophes are the codimension 1 surfaces of the control parameter space (α, β, ξ) for which a polynomial admits a double root. We find that (4.5) admits multiple solutions for the domain of the control space lying between the zero circulation surface and the suitable part of the surface defined by

$$4\alpha\xi^4 - \xi^4 - 2\beta\xi^3 - 8\alpha\xi^3 + 2\xi^3 - \beta^2\xi^2 + 20\alpha\beta\xi^2 - 2\beta\xi^2 + 8\alpha^2\xi^2 + 2\alpha\xi^2 - \xi^2 - 8\beta^2\xi - 38\alpha\beta\xi + 8\beta\xi - 8\alpha^2\xi + 2\alpha\xi - 4\beta^3 + 12\alpha\beta^2 + 8\xi^2 - 12\alpha^2\beta + 20\alpha\beta - 4\beta + 4\alpha^3 - \alpha^2 = 0.$$

This analytical calculation was made with the use of the symbolic manipulation software Macsyma.

C. Stability analysis of the asymptotic three-box model

We want to investigate the stability of the nine equilibria $(\Psi_s, \Psi_n) = (\alpha X^{(m_s)}, \alpha X^{(m_n)})$ for the dynamical system (5.2), where the quantities $X^{(m_i)}$, $i \in \{s, n\}$, are chosen among the solutions $X^{(1)} = (1 - \sqrt{1 + 4Y})/2$, $X^{(2)} = (1 - \sqrt{1 - 4Y})/2$ and $X^{(3)} = (1 + \sqrt{1 - 4Y})/2$ of the implicit equation $Y = |X|(1 - X)$ (see Figure 9a), provided that the quantity $Y = \beta/(\alpha|\alpha|)$ is chosen in the interval $[0, 1/4]$.

We call $\Lambda_i = -\eta_i(\alpha - 2\Sigma_i) = \alpha\eta_i(1 - 2X^{(m_i)})$, $i \in \{s, n\}$, where η_i is the sign of Ψ_i. The stability of an equilibrium (Ψ_s, Ψ_n) is given by the real part of the roots of the characteristic equation

$$4\lambda^2 - 3(\Lambda_s + \Lambda_n)\lambda + 2\Lambda_s\Lambda_n = 0,$$

which is obtained for infinitesimal perturbations growing like $\exp \lambda t$ in a linearization of (5.2) around the equilibrium.

For the three symmetric equilibria $(\alpha X^{(1)}, \alpha X^{(1)})$, $(\alpha X^{(2)}, \alpha X^{(2)})$ and $(\alpha X^{(3)}, \alpha X^{(3)})$ we have $\Lambda_s = \Lambda_n = \Lambda$. One eigenvalue, $\lambda = \Lambda$, comes from the symmetric problem (i.e., the two-box model); it takes the value $\Lambda^{(1)} = -\alpha\sqrt{1 + 4Y}$, $\Lambda^{(2)} = \alpha\sqrt{1 - 4Y}$ and $\Lambda^{(3)} = -\alpha\sqrt{1 - 4Y}$ respectively. The other eigenvalue, $\lambda = \Lambda/2$, corresponds to a purely antisymmetric mode. Both eigenvalues have the same sign and we conclude that SA $= (\alpha X^{(1)}, \alpha X^{(1)})$ and TH $= (\alpha X^{(3)}, \alpha X^{(3)})$ are the only stable symmetric equilibria.

For the six asymmetric equilibria $(\alpha X^{(3)}, \alpha X^{(1)})$, $(\alpha X^{(3)}, \alpha X^{(2)})$, $(\alpha X^{(2)}, \alpha X^{(1)})$ and symmetric counterparts, we can show that only PP $= (\alpha X^{(3)}, \alpha X^{(1)})$ or $(\alpha X^{(1)}, \alpha X^{(3)})$ are stable. Indeed, the eigenvalues are real as shown by the positivity of the

charcateristic equation discriminants: $\alpha^2(18 - 14\sqrt{1 - 16Y^2})$, $32\alpha^2(1 - 4Y)$ and $\alpha^2(18 + 14\sqrt{1 - 16Y^2})$ respectively for the three asymmetric equilibria and their symmetric counterparts. For the two PP equilibria the eigenvalues have a positive product, $(\alpha^2\sqrt{1 - 16Y^2})/2$, and the sum, $-3\alpha(\sqrt{1 + 4Y} + \sqrt{1 - 4Y})/4$ is negative, which implies that their are both negative: the two PP equilibria are thus stable. For the other asymmetric equilibria the eigenvalues have a negative product, $-\alpha^2(1 - 4Y)/2$ and $-(\alpha^2\sqrt{1 - 16Y^2})/2$ respectively, which implies that one of them is positive: these equilibria are unstable.

This stability analysis can be summed up by saying that an equilibrium is unstable whenever it contains the unstable half-domain solution $X^{(2)}$. The "superposition principle" requires that each half of the domain contain one of the stable half-domain equilibria in order to build a stable equilibria on the full domain.

D. 3 × 2 box model equations

With the choice of units described in Section 6.1, the 3 × 2 box model equations (see Figure 10) are

$$\dot{T}_s^u = -\tfrac{1}{2}\alpha - T_s^u + \tfrac{1}{2}(\Psi_s^- + \kappa_v)(T_s^l - T_s^u) + \tfrac{1}{2}(\Psi_s^+ + \kappa_h)(T_e^u - T_s^u),$$

$$\dot{T}_e^u = \tfrac{1}{2}\alpha - T_e^u + \tfrac{1}{2}[\tfrac{1}{2}(\Psi_s^+ + \Psi_n^+) + \kappa_v](T_e^l - T_e^u) + \tfrac{1}{4}(\Psi_s^- + \kappa_h)(T_s^u - T_e^u) + \tfrac{1}{4}(\Psi_n^- + \kappa_h)(T_n^u - T_e^u),$$

$$\dot{T}_n^u = -\tfrac{1}{2}\alpha - T_n^u + \tfrac{1}{2}(\Psi_n^- + \kappa_v)(T_n^l - T_n^u) + \tfrac{1}{2}(\Psi_n^+ + \kappa_h)(T_e^u - T_n^u),$$

$$\delta\dot{T}_s^l = \tfrac{1}{2}(\Psi_s^+ + \kappa_v)(T_s^u - T_s^l) + \tfrac{1}{2}(\Psi_s^- + \delta\kappa_h)(T_e^l - T_s^l),$$

$$\delta\dot{T}_e^l = \tfrac{1}{2}[\tfrac{1}{2}(\Psi_s^- + \Psi_n^-) + \kappa_v](T_e^u - T_e^l) + \tfrac{1}{4}(\Psi_s^+ + \delta\kappa_h)(T_s^l - T_e^l) + \tfrac{1}{4}(\Psi_n^+ + \delta\kappa_h)(T_n^l - T_e^l),$$

$$\delta\dot{T}_n^l = \tfrac{1}{2}(\Psi_n^+ + \kappa_v)(T_n^u - T_n^l) + \tfrac{1}{2}(\Psi_n^- + \delta\kappa_h)(T_e^l - T_n^l),$$

and

$$\dot{S}_s^u = -\tfrac{1}{2}\beta - \xi S_s^u + \tfrac{1}{2}(\Psi_s^- + \kappa_v)(S_s^l - S_s^u) + \tfrac{1}{2}(\Psi_s^+ + \kappa_h)(S_e^u - S_s^u),$$

$$\dot{S}_e^u = \tfrac{1}{2}\beta - \xi S_s^u + \tfrac{1}{2}[\tfrac{1}{2}(\Psi_s^+ + \Psi_n^+) + \kappa_v](S_e^l - S_e^u) + \tfrac{1}{4}(\Psi_s^- + \kappa_h)(S_s^u - S_e^u) + \tfrac{1}{4}(\Psi_n^- + \kappa_h)(S_n^u - S_e^u),$$

$$\dot{S}_n^u = -\tfrac{1}{2}\beta - \xi S_n^u + \tfrac{1}{2}(\Psi_n^- + \kappa_v)(S_n^l - S_n^u) + \tfrac{1}{2}(\Psi_n^+ + \kappa_h)(S_e^u - S_n^u),$$

$$\delta\dot{S}_s^l = \tfrac{1}{2}(\Psi_s^+ + \kappa_v)(S_s^u - S_s^l) + \tfrac{1}{2}(\Psi_s^- + \delta\kappa_h)(S_e^l - S_s^l),$$

$$\delta \dot{S}_e^l = \tfrac{1}{2}[\tfrac{1}{2}(\Psi_s^- + \Psi_n^-) + \kappa_v](S_e^u - S_e^l) + \tfrac{1}{4}(\Psi_s^+ + \delta\kappa_h)(S_s^l - S_e^l)$$
$$\tfrac{1}{4}(\Psi_n^+ + \delta\kappa_h)(S_n^l - S_e^l),$$

$$\delta \dot{S}_n^l = \tfrac{1}{2}(\Psi_n^+ + \kappa_v)(S_n^u - S_n^l) + \tfrac{1}{2}(\Psi_n^- + \delta\kappa_h)(S_e^l - S_n^l),$$

where the circulations are assumed to follow capillary laws with the circulation intensity proportional to the depth-integrated lateral buoyancy difference,

$$\Psi_s = \frac{1}{1+\delta}[(T_e^u - T_s^u) - (S_e^u - S_s^u)] + \frac{\delta}{1+\delta}[(T_e^l - T_s^l) - (S_e^l - S_s^l)],$$

$$\Psi_n = \frac{1}{1+\delta}[(T_e^u - T_n^u) - (S_e^u - S_n^u)] + \frac{\delta}{1+\delta}[(T_e^l - T_n^l) - (S_e^l - S_n^l)].$$