Tellus (2004), 56A, 362–370
Printed in UK. All rights reserved

TELLUS

A systematic approach to determine thresholds of the ocean's thermohaline circulation

By HENK A. DIJKSTRA[1]*, LIANKE TE RAA[1] and WILBERT WEIJER[2], [1]*Institute for Marine and Atmospheric Research Utrecht, Department of Physics and Astronomy, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands;* [2]*Scripps Institute of Oceanography, University of California at San Diego, La Jolla, USA*

(Manuscript received 2 June 2003; in final form 29 October 2003)

ABSTRACT

A systematic approach is proposed to determine thresholds in freshwater flux perturbations related to abrupt changes in the ocean's thermohaline circulation. The typical problem considered is the response of a thermohaline driven flow to a localized change, of specified strength and duration, in the surface freshwater flux. The initial transient response due to the freshwater anomaly is considered as a finite amplitude perturbation. An estimate of this response can be obtained by using ideas from dynamical systems theory. Central quantity to determine whether such a perturbation leads to instability (i.e. a 'collapsed' state) is the sign of the tendency of a specific energy functional. The approach is first illustrated with a simple box model and then shown to give good results in a global ocean general circulation model.

1. Introduction

The oxygen isotope records from the Greenland ice cores have indicated that the climate system may undergo relatively rapid changes. One of the most famous events is the Younger Dryas cooling that occurred at about 12 ka during the transition from the last glacial period to the Holocene. Another example is the abrupt cooling event that occurred about 8000 yr ago, with an estimated temperature difference of 3–6 °C over Greenland (Alley et al. 1997).

One of the potential physical mechanisms of these abrupt cooling events is the change in the global ocean circulation due to changes in salinity in the North Atlantic (Broecker 1997). The present global thermohaline circulation is imagined as an 'Ocean Conveyor', which is driven by lateral buoyancy gradients and localized sites of deep water formation, with a prominent role for the North Atlantic (Schmitz 1995). There is a net sinking of deep water of about 15 Sv in the North Atlantic, which is associated with the formation of North Atlantic Deep Water (NADW). The NADW can be traced down to 30 °S in the South Atlantic and connects more southwards to water masses from the Southern Ocean.

The Atlantic part of the global ocean circulation is sensitive to the amplitude and spatial pattern of the surface buoyancy forcing. The reason is that small buoyancy perturbations can be amplified through feedback mechanisms, in particular the salt-advection feedback (Stommel 1961; Rooth 1982). A hypothesis on the cause of the Younger Dryas cooling event is that it started with a catastrophic input of freshwater through diversion of Lake Agassiz from the Mississippi basin to the St Lawrence drainage (Rooth 1982). This slowed down the Atlantic branch of the global ocean circulation with a consequent decrease in northward heat transport. A similar mechanism of sudden freshwater input may have caused the Holocene cooling event about 8000 yr ago. A central issue in this hypothesis is whether the changes in freshwater input in the past have had sufficient amplitude to cause the reconstructed temperature changes on Greenland. This is one of the reasons why the sensitivity of the ocean's global circulation has become one of the important issues in climate research.

In many models of the global ocean circulation, it appears that several equilibrium states may exist under similar forcing conditions. When the present 'Conveyor' is subjected to a quasi-steady freshwater flux perturbation in the North Atlantic, eventually the circulation may collapse. In the resulting equilibrium state, there is deep water formation in the Southern Ocean instead of in the North Atlantic and the formation of the NADW has ceased (Stocker et al. 1992; Rahmstorf 1995; Manabe and Stouffer 1999).

Schiller et al. (1997) consider the response of the coupled global LSG/ECHAM3 model to freshwater input into the Labrador Sea. The freshwater input is increased linearly for 250 yr to a maximum of 0.625 Sv and then linearly reduced (over the same time period) to zero. Initially, the NADW formation was suppressed completely and a reversed Atlantic meridional

*Corresponding author. Present address: Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523-1371.
e-mail: dijkstra@atmos.colostate.edu

overturning circulation was found. However, after the freshwater input stopped, the original overturning circulation recovered and no permanent shut down occurred.

Manabe and Stouffer (1995) investigate the effect of the location of the freshwater input on the ocean circulation in the coupled Geophysical Fluid Dynamics Laboratory (GFDL) climate model. When 0.1 Sv is discharged over an area in the northern North Atlantic for a period of 500 yr, the overturning decreases from 18 Sv to about 4 Sv. Similar to the results in Schiller et al. (1997), the circulation recovers in about 250 yr after the anomalous forcing is removed. The circulation is less sensitive to anomalous freshwater input in the subtropical North Atlantic, decreasing only by about 4 Sv.

More recently, Vellinga et al. (2002) have investigated the response of a coupled ocean–atmosphere model (HadCM3) to a sudden negative change in surface salinity in the northern North Atlantic. The overturning circulation is strongly reduced initially, but it recovers after about 120 yr. The salt transport by the subtropical gyre appears a crucial factor in the recovery process because it is able to restore the salt deficit caused by the initial perturbation.

While there are only a few simulations with fully coupled general circulation models, many more results have been obtained with so-called intermediate complexity climate models. Fanning and Weaver (1997) impose different sequences of run-off at four different locations in the North Atlantic, based on reconstructed conditions during the Younger Dryas. It appears that the circulation is more vulnerable to a sequence of freshwater inputs than to a single one and a fully collapsed state can be maintained even if the anomalous forcing is removed.

In more idealized models, for example climate models where the ocean model has a near two-dimensional dynamics (Stocker et al. 1992; Ganopolsky et al. 2001), collapsed states as a result of anomalous freshwater inputs have also been frequently found. From these results, it appears that there exist thresholds in the input of freshwater, such that the circulation totally collapses. These thresholds depend in a complicated way on ocean model parameters.

In a system that potentially has multiple equilibria, the response to a finite amplitude perturbation induced by anomalous forcing conditions is an intricate problem. In this paper, we develop a systematic approach that aims to determine thresholds of the thermohaline circulation with respect to freshwater perturbations. These boundaries separate regimes where the thermohaline circulation recovers or collapses after a finite-time perturbation in the freshwater flux. The approach is based on elementary ideas of bifurcation theory. It is illustrated first with a simple box model and then applied to a global ocean circulation model.

2. Theory

The starting point is an equilibrium state of the ocean circulation, say computed with a particular model. Many large-scale time-mean equilibrium flows can be idealized as steady-state flows. In fact, model–model comparison studies indicate that the large-scale time-mean states closely correspond to the steady states (Dijkstra 2000). The idealization of steady states allows a more rigorous analysis of the parameter dependence of these flows. We will refer to the problem of how the system responds to the temporary change of the freshwater flux as the 'thermohaline pulse response' (TPR) problem; it is specified in more detail below.

2.1. Thermohaline pulse response problem

The steady-state ocean flow is maintained by wind-stress forcing and a surface heat and freshwater flux; we assume that these forcing fields are given. The spatial pattern of the freshwater flux field is indicated by $\bar{F}_S(\mathbf{x})$ and its amplitude by γF_0. Because of salt conservation, the following integral condition

$$\int_S \gamma F_0 \bar{F}_S \mathrm{d}^2x = 0 \tag{1}$$

holds where S is the ocean–atmosphere surface.

Assume there is a freshwater flux anomaly of a certain amplitude δ_F (ms $^{-1}$) and given pattern $\tilde{F}_S$ that is applied for a time t_m (s). If we idealize the time dependence as a block function $B(t;t_\mathrm{m})$, then the total freshwater flux can be written as

$$F_S(\mathbf{x}, t; t_\mathrm{m}) = \gamma F_0 \bar{F}_S + \delta_F \tilde{F}_S(\mathbf{x}) B(t; t_\mathrm{m}) - Q(t) \tag{2a}$$

$$B(t; t_\mathrm{m}) = H(t) - H(t - t_\mathrm{m}) \tag{2b}$$

where H is the Heaviside function and $Q(t)$ is determined from

$$\int_S F_S \mathrm{d}^2x = 0 \Rightarrow Q(t) = B(t; t_\mathrm{m}) \frac{\delta_F}{|S|} \int_S \tilde{F}_S(\mathbf{x}) \mathrm{d}^2x \tag{3}$$

and $|S|$ is the surface area. The constraint (3) ensures overall salt conservation. In case a land-based freshwater reservoir is suddenly drained into the ocean, such as perhaps during the Younger Dryas, a net input of freshwater occurs; it can be represented by adjusting the value of Q.

The TPR problem is concerned with the response of the ocean circulation under different conditions determined by the spatial pattern $\tilde{F}_S$ and the values of the parameters δ_F and t_m. This response will also depend on the ocean model parameter values. When there is only a single steady state for the values of the parameters chosen, say indicated by A_1, this state must be globally stable. This means that the ocean circulation is affected by the anomalous freshwater forcing, but as soon as the forcing has disappeared (i.e. for $t > t_\mathrm{m}$), the ocean flow will recover to the original state A_1.

However, when there are more stable equilibria, say states A_1 and A_2, the situation is different. When the anomalous forcing is applied to the initial state A_1, the state that is reached at $= t_\mathrm{m}$ may evolve towards the state A_2 (for $t > t_\mathrm{m}$). In this way, a transition takes place that can be viewed as a finite amplitude instability of state A_1. The central issue in the TPR problem is how to determine whether this instability will occur and how the

critical boundary depends on the time t_m and the perturbation strength δ_F.

2.2. Finite amplitude perturbations

We approach the TPR problem by using the parameter dependence of the equilibria. Consider in Fig. 1a typical bifurcation diagram associated with thermohaline flows (Dijkstra 2000) in one of the control parameters, say λ. Here the diagram is computed from the Stommel (1961) two-box model, explained in more detail in the next section. In other models, this parameter may indicate the viscosity of the flow or the strength of the surface freshwater flux.

Assume that the bifurcation diagram for $\delta_F = 0$ is represented by the curve C_1 in Fig. 1. In addition, we can also consider the same diagram for the case $t_m \to \infty$ when the forcing is given by

$$\hat{F}_S(\mathbf{x}) = \gamma F_0 \bar{F}_S + \delta_F \tilde{F}_S(\mathbf{x}) - Q \tag{4}$$

where Q is a single scalar determined from eq. (3) with $B(t;t_m) = 1$. Because this forcing is steady, we can also calculate the bifurcation diagram for this case and a possible outcome is plotted as the curve C_2 in Fig. 1.

Next, we define the stable state A_1 on curve C_1 (for a particular fixed value of λ) as the basic state. On this basic state we apply the pulse forcing with amplitude δ_F over a time t_m. The model trajectory can be quite complicated, but eventually a state I is reached at $t = t_m$; a possible outcome is indicated in Fig. 1. Note that for $t_m \to \infty$ – a permanent freshwater anomaly – the stable state B on the curve C_2 is reached.

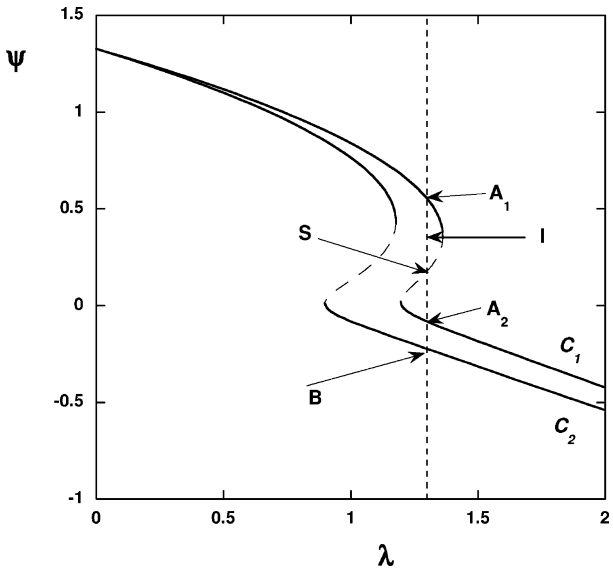

Fig 1. Typical bifurcation diagram of the thermohaline circulation in which the strength of the overturning circulation Ψ is plotted versus a measure of the freshwater flux λ (Stommel 1961). The two curves are for different ratios of the surface restoring time-scales of salt and heat.

As a sidestep, a rough approximation of the state I can be obtained in case the linear stability problem for the stable state B can be solved. In that case, a normal mode approach provides the eigenvalues and eigenvectors that control the evolution of infinitesimally small perturbations on state B. Let the least stable mode have a complex growth rate σ given by

$$\sigma = \sigma_r \pm i\sigma_i \tag{5}$$

with $\sigma_r < 0$. The latter determines the attraction to state B along part of its stable manifold. A rough approximation of a trajectory that connects state A_1 and B is then

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{u}_{A_1} + (\mathbf{u}_B - \mathbf{u}_{A_1})(1 - e^{\sigma_r t}\cos\sigma_i t) \tag{6}$$

where $\mathbf{u}$ denotes the total state vector of the flow. Indeed, for $t = 0$, this trajectory starts from state A_1, it is attracted to state B and reaches it for $t \to \infty$. The vector $\mathbf{u}_I$ of the state I is then approximated by taking $t = t_m$ in eq. (6).

For $t > t_m$, the forcing is again $\bar{F}_S$ and only the equilibria on the curve C_1 are relevant for the evolution of state I. For the chosen value of λ, these are the original states A_1 and A_2 as well as the unstable state S (Fig. 1). The evolution from state I may be complex because the attraction domains of the states A_1 and A_2 can have a complicated (e.g. fractal) structure. However, one can attempt to base the attraction domain on a single scalar quantity by referring to basin-integrated quantities.

2.3. Energy principles

In many large-scale ocean flows, the effect of inertia can be neglected and the energy of the flow is dominated by the available potential energy, which is linked to the density anomalies $\tilde{\rho}$ of the flow. This motivates us to define an energy functional $\mathcal{E}$ defined as (Weijer and Dijkstra 2001)

$$\mathcal{E} = \langle \tilde{\rho}^2 \rangle \tag{7}$$

where $\langle . \rangle$ indicates integration over the flow domain. If there are two attracting states (i.e. A_1 and A_2 in Fig. 1), then two density anomalies $\tilde{\rho}_i = \rho_{A_i} - \rho$, $i = 1, 2$ can be considered, with corresponding energy functionals $\mathcal{E}_i$. In addition, the energy $\mathcal{E}_s$ with respect to the state S can be defined, with $\tilde{\rho}_s = \rho_S - \rho$.

To determine whether state I will evolve back to A_1 or will change to state A_2 for $t > t_m$, one would need to have information on the separatrix bounding the basin of attractors of both equilibria. However, information is difficult to obtain, other than that the unstable equilibrium S is necessarily on this separatrix. Intuitively, however, one expects that the evolution of the trajectory will be associated with changes in energy tendencies of the above-defined functionals with a special role for the energy with respect to the unstable state. We will demonstrate below that in the relevant situations of the thermohaline circulation, the energy functionals indeed may be useful as indicators of instability.

The TPR problem is then tackled by the following computational procedure.

(i) For a chosen value of δ_F, calculate the bifurcation diagrams in a relevant control parameter λ of the flow for both the original forcing ($\bar{F}_S$) and the 'permanent' freshwater forcing ($\hat{F}_S$).

(ii) Choose a fixed value of λ and label one of the equilibria (similar to A_1) as initial state. Apply the forcing $\hat{F}_S$ (with $t_m \to \infty$) and compute the transient flow until equilibrium. This trajectory will end up in a stable state on the bifurcation curve for $\hat{F}_S$ (similar to state B).

(iii) Compute the energy functionals (with respect to the equilibria for $\bar{F}_S$) along the trajectory computed in the previous step. The value of t_m for which a transition will occur, say t_m^*, is located near the point where the tendency of $\mathcal{E}_s$ changes sign.

Consequently, when $t_m < t_m^*$ the trajectory will approach the original state for $t > t_m$, whereas a different state will be reached when $t_m > t_m^*$.

3. Application to a simple box model

To illustrate the approach, the theory is first applied to a two-box model (Stommel 1961) of the thermohaline circulation. This model is known to have multiple equilibria of which the structure in parameter space is easily computed. The box model formulation is presented in many textbooks (we use the notation in chapter 3 of Dijkstra 2000) and the dimensionless equations become

$$\frac{dT}{dt} = \eta_1 - T(1 + M(T - S)) \tag{8a}$$

$$\frac{dS}{dt} = \eta_2 - S(\eta_3 + M(T - S)) \tag{8b}$$

where $T = T_e - T_p$, $S = S_e - S_p$ are the scaled temperature and salinity difference between the equatorial and polar box, and $\Psi = T - S$ is the dimensionless flow rate. The function M indicates the modulus function. Three parameters appear in eq. (8): the parameters η_1 and η_2 measure the strength of the thermal forcing and of the freshwater forcing, respectively, and η_3 is the ratio of the relaxation times of temperature and salinity.

As there is no spatial structure in the forcing, we will change another parameter in addition to the freshwater flux amplitude η_2. In the example below, we choose to fix $\eta_1 = 3.0$ and change η_3. For a 'standard' value of $\eta_3 = 0.4$, the bifurcation diagram is plotted in Fig. 2 as the curve with the '0.4' label. As a norm of the solution, the equator to pole (dimensionless) temperature (Fig. 2a) and salinity (Fig. 2b) difference are taken. The saddle node bifurcations occur at $\eta_2 = 1.19$ and $\eta_2 = 1.36$ and bound the region of multiple equilibria. If we focus on a value $\eta_2 = 1.3$, then there are two stable states. The state A_1 is called a thermally driven state (or TH state) with $\Psi > 0$ and the state A_2 is called a salinity-driven state (or SA state) with $\Psi < 0$. Both are indicated in Fig. 2 and also the unstable state S is labeled.

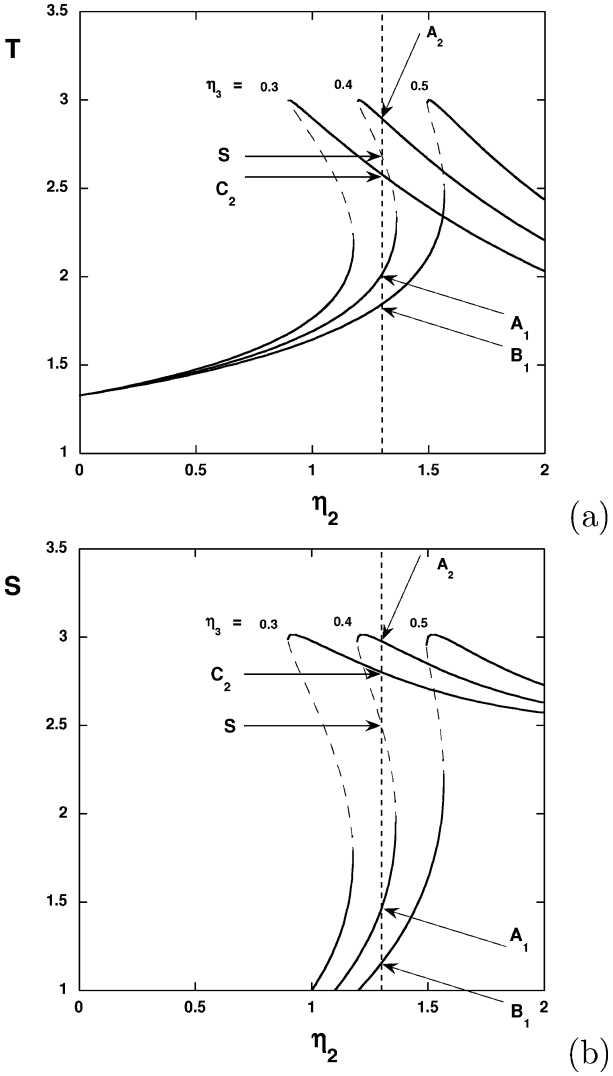

Fig 2. (a) Temperature and (b) salinity along the steady states of the Stommel two-box model bifurcation diagram, with $\eta_1 = 3.0$. Stable steady states are indicated by solid curves, while curves of unstable steady states are dashed.

The bifurcation diagram for $\eta_3 = 0.5$ is also plotted in both panels of Fig. 2. Although the bifurcation diagram is qualitatively similar to that of $\eta_3 = 0.4$, there is only one steady state for $\eta_2 = 1.3$; this is a TH state, which is denoted B_1. Hence, a trajectory starting at A_1, evolving under conditions represented by $\eta_3 = 0.5$, will be attracted to B_1. Because this state is also thermally driven, the overturning is only slightly modified with respect to that of A_1. Similarly by changing η_3 to 0.3, the bifurcation diagram does not change qualitatively (Fig. 2). For $\eta_2 = 1.3$, there is only one steady (SA) state, which is labeled C_2. Now a trajectory starting in A_1 will be attracted to C_2, a state that is qualitatively different from A_1 and has equatorial sinking; a collapse will occur.

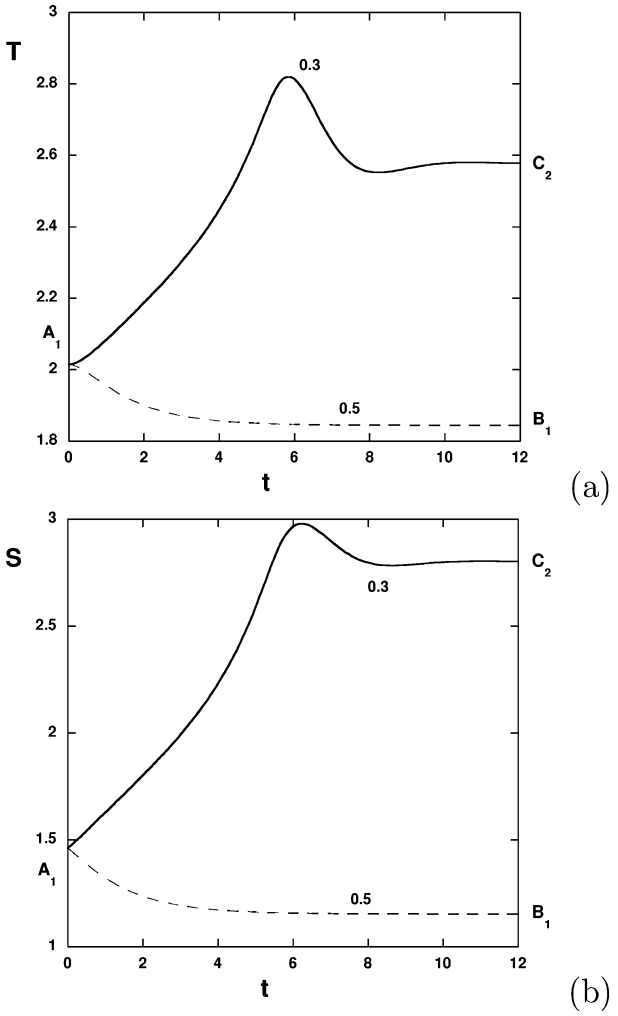

Fig 3. Trajectories from the initial condition A_1 for two different values of η_3 ($\eta_3 = 0.3$ and $\eta_3 = 0.5$): (a) temperature and (b) salinity.

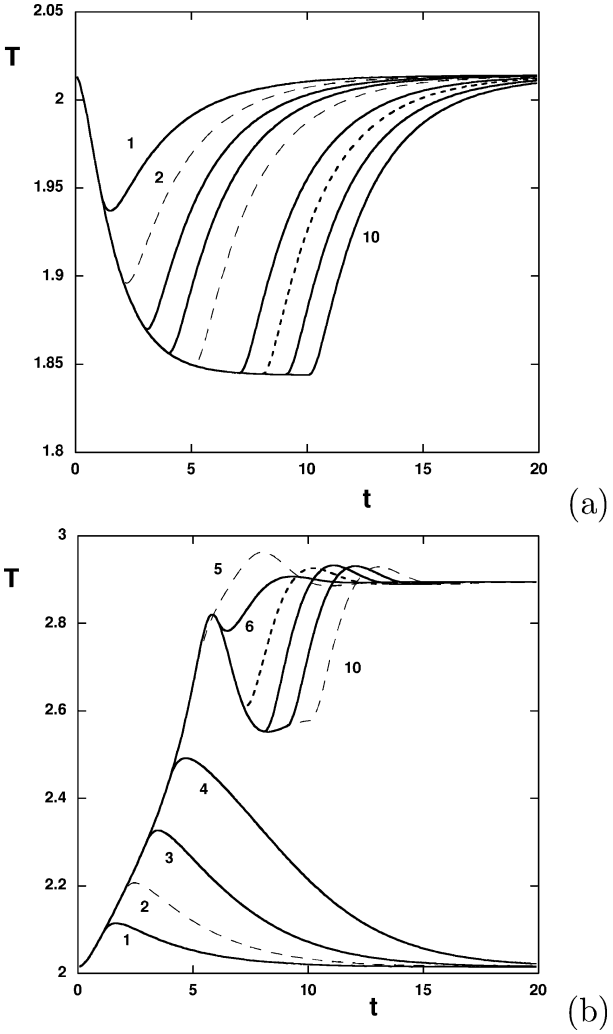

Fig 4. Trajectories for (a) $\eta_1 = 3.0$, $\eta_2 = 1.3$ and $\eta_3 = 0.5$ and (b) $\eta_1 = 3.0$, $\eta_2 = 1.3$ and $\eta_3 = 0.3$. The different curves are for different dimensionless times t_m.

Both types of trajectories are seen in Fig. 3: one under conditions $\eta_3 = 0.5$ (going from A_1 to B_1) and one under conditions $\eta_3 = 0.3$ (going from A_1 to C_2). From these trajectories, the attraction properties of state B_1 and C_2 can be determined. The time-scale of approach to the equilibrium is in agreement with the eigenvalues computed at both states from a linear stability analysis of state B_1 and C_2. In fact, the least stable mode of state C_2 is an oscillatory mode, which can be seen as the oscillatory behavior of the trajectory.

We now apply the pulse forcing to the system with $\eta_3 = 0.4$ and $\eta_2 = 1.3$. From $t = 0$ up to $t = t_m$, the parameter η_3 is changed to either 0.3 or 0.5 and the initial development from state A_1 is computed. After $t = t_m$, the parameter η_3 is set back to its original value of 0.4. This effectively mimics the TPR problem, because a freshwater 'forcing' is introduced in eq. (8) over the time interval $[0, t_m]$.

In Fig. 4a, the temperature along the resulting trajectories is plotted for different labeled values of t_m for the case that η_3 is set temporarily to 0.5. The lower 'envelope' of the curves corresponds to the trajectory $A_1 \rightarrow B_1$ in Fig. 3a. The curves look qualitatively very similar to those obtained in complex climate models after a freshwater anomaly has been imposed (Manabe and Stouffer 1995). As soon as the anomalous forcing is removed at time t_m, each trajectory approaches the state A_1 for $t \rightarrow \infty$. This indicates that the anomaly is not able to 'lure' the trajectory into the attraction domain of the state A_2.

As could be expected, the situation is rather different for the case $\eta_3 = 0.3$. Up to a time $t_m \leq t_m^*$, the trajectories return to state A_1 as the 'forcing' is removed. However, for $t_m \geq t_m^*$ a critical threshold is passed and the trajectories are attracted to state A_2; a collapse of the thermally driven circulation occurs. The signature of the oscillatory attraction of state A_2 is here clearly visible.

From Fig. 4b, we can conclude that $t_{\mathrm{m}}^{*} \in [4, 5]$ and by repeated integrations it is found that $t_{\mathrm{m}}^{*} = 4.801$. However, we would like to have an estimate of this value by using the tendencies of the energy functionals only along the trajectory $A_1 \rightarrow C_2$ in Fig. 3a.

Because the stream function is linearly proportional to the density, we define the dimensionless energy functional as

$$\mathcal{E} = \frac{1}{2}(\tilde{T} - \tilde{S})^2 \tag{9}$$

where $\tilde{T} = T - \bar{T}$, $\tilde{S} = S - \bar{S}$, and $(\bar{T}, \bar{S})$ is one of the steady states A_1, A_2 or S. In most models, these energy functionals can be directly computed once the steady states are known.

For the special case of the two-box model, we can actually calculate the tendencies of the energy functionals exactly. The evolution equations for $\tilde{T}$ and $\tilde{S}$ are

$$\frac{\mathrm{d}\tilde{T}}{\mathrm{d}t} = \eta_1 - (\bar{T} + \tilde{T})(1 + M(\bar{\Psi} + \tilde{\Psi})) \tag{10a}$$

$$\frac{\mathrm{d}\tilde{S}}{\mathrm{d}t} = \eta_2 - (\bar{S} + \tilde{S})(\eta_3 + M(\bar{\Psi} + \tilde{\Psi})). \tag{10b}$$

By multiplying both eq. (10a) and eq. (10b) with $\tilde{T}$ and $\tilde{S}$, the equation for $\mathrm{d}\mathcal{E}/\mathrm{d}t$ becomes

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = \mathcal{I} - \mathcal{D}_1 - \mathcal{D}_2 \tag{11}$$

where $\mathcal{I}$ is an energy production term and $\mathcal{D}_i$ represent dissipation effects (expressions not shown).

In most other models, we cannot compute the tendencies explicitly and hence we show for the box model only the energy functionals computed directly from eq. (9). Along the trajectories in Fig. 3, the energy functionals with respect to the different equilibria A_1, A_2 and S are plotted for the forcing with a temporary value of $\eta_3 = 0.3$ (Fig. 5a) and $\eta_3 = 0.5$ (Fig. 5b). In both cases, the trajectory is drifting away from A_1 as the energy with respect to this equilibrium increases. For $\eta_3 = 0.5$, the energies with respect to A_2 and S both increase and no transition behavior is expected. However, for $\eta_3 = 0.3$, the tendency of $\mathcal{E}_s$ changes sign at $t = 4.35$. This gives an indication that, for $t > 4.35$, the unstable steady state has been passed. This value is close to the value of t_{m}^{*} as determined directly from the model.

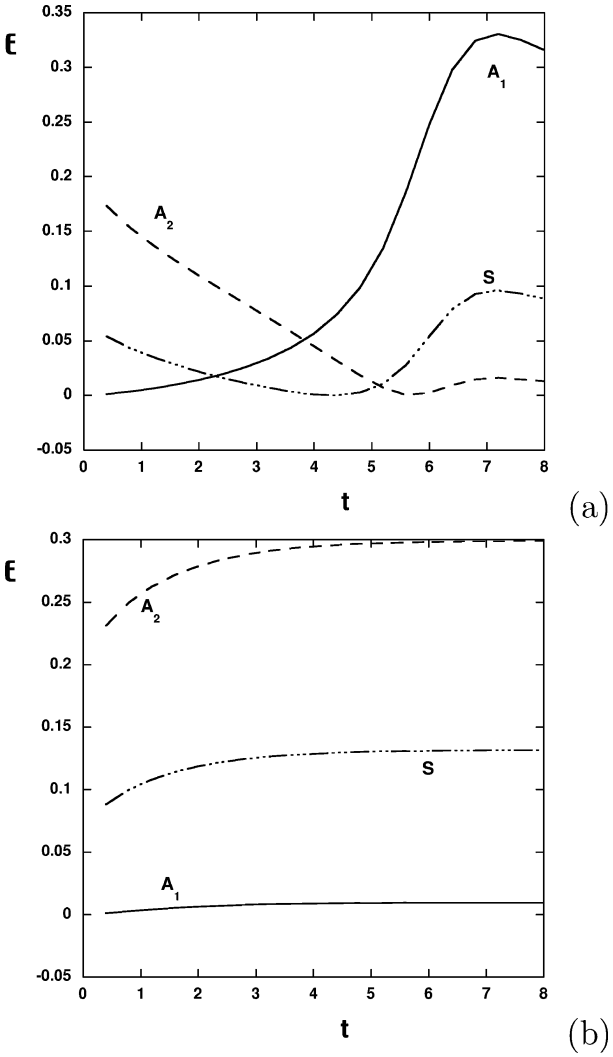

Fig 5. Energy tendencies for the different steady states A_1, A_2 and S along the trajectories in Fig. 3: (a) $\eta_3 = 0.3$; (b) $\eta_3 = 0.5$.

4. A global ocean–atmosphere model

Recently, it has become possible to compute bifurcation diagrams for ocean general circulation models. In Weijer et al. (2003), this was done for a model of the global ocean that is coupled to an energy balance model of the atmosphere. The governing equations of the ocean model are the hydrostatic, primitive equations in spherical coordinates on a global domain, which includes full continental geometry as well as bottom topography. The horizontal resolution is about 4°, with a 96 × 38 grid on a domain $[0, 360] \times [-85.5, 85.5]$. The grid has 12 levels in the vertical and is non-equidistant with the most upper (lowest) layer having a thickness of 50 m (1000 m), respectively.

Starting from the trivial state (at zero solar forcing, no freshwater flux and no wind stress) first an equilibrium state is determined under the annual-mean wind stress as in Trenberth et al. (1989), an analytical form of the solar forcing and the Levitus surface salinity distribution (Levitus et al. 1994). The freshwater flux of this reference solution is diagnosed and subsequently used as part of the surface buoyancy forcing. Next, a steady freshwater flux perturbation with an amplitude γ_{p} (in Sv) is applied over a relatively small domain, say $\mathcal{D}$, in the North Atlantic ($\mathcal{D} = \{(\phi, \theta) \in [300, 336] \times [54, 66]\}$).

In Weijer et al. (2003), bifurcation diagrams were computed in the two-parameter plane spanned by γ and γ_{p}. The dimensionless parameter γ indicates the strength of the diagnosed

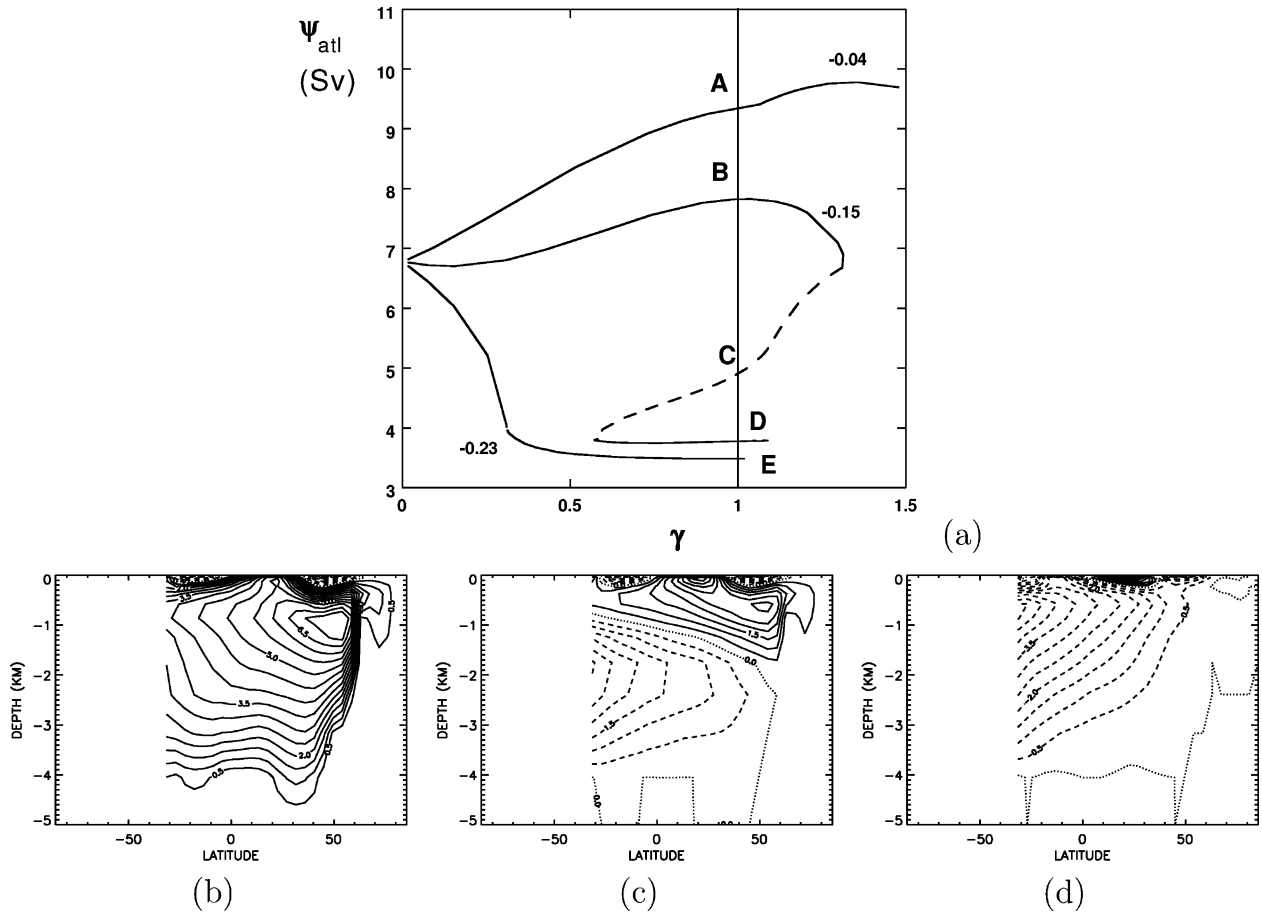

Fig 6. (a) Bifurcation diagrams for several values of γ_p as curves of the maximum Atlantic meridional overturning versus the strength of the diagnosed freshwater flux γ. Again stable states are indicated by solid curves and dashed curves indicate unstable states. (b) Pattern of the meridional stream function in the Atlantic for the state B in (a). (c) Similar to (b) but for state C. (d) Similar to (b) but for state D.

freshwater flux. The case $\gamma = 1$ corresponds to the freshwater flux that is needed to maintain the Levitus surface salinity field. The emergence of the multiple equilibrium regime with changing freshwater flux can be seen by plotting bifurcation diagrams for several fixed values of γ_p (Fig. 6a) using γ as control parameter. For $\gamma_p = -0.04$ Sv, the overturning increases with γ, while no multiple equilibria occur. For $\gamma_p = -0.15$ Sv, however, two saddle-node bifurcations appear and a regime of multiple equilibria exists. The Atlantic meridional overturning stream functions of the solutions at the locations B, C and D ($\gamma = 1$) are plotted in Figs 6b–d , respectively. State B still has substantial northward sinking, while in state D northern sinking is absent. In the unstable state C, the northern sinking is substantially reduced but southern sinking is still relatively weak. For $\gamma_p = -0.23$ Sv, the overturning decreases monotonically with γ and there are no multiple equilibria. Only the collapsed state E exists for $\gamma = 1$, with a meridional overturning pattern similar to state D (Fig. 6d).

To illustrate the approach to the TPR problem in this model, we focus on the bifurcation diagram for $\gamma_p = -0.15$ and fix $\gamma = 1$. We take state B as the initial state and want to determine the behavior of the system once an anomalous freshwater flux forcing is applied (over the domain $\mathcal{D}$) to this state. The bifurcation diagram then indicates that state B may collapse to state D (Fig. 6). We choose $\delta_F = -0.08$ Sv such that for a permanent freshwater perturbation the bifurcation diagram is that for $\gamma_p = -0.23$ Sv. Starting at state B, we change γ_p from -0.15 to -0.23 at $t = 0$ and monitor the evolution of the flow. This is shown as the solid curve in a plot (Fig. 7a) of the Atlantic overturning versus time (in yr). Consistent with the bifurcation diagrams in Fig. 6, after about 1000 yr the state E is approached.

Along this trajectory, we monitor the energy functionals with respect to the states B, C and D (Fig. 7b). The energy $\mathcal{E}_B$ increases monotonically, whereas the tendency of the energy $\mathcal{E}_C$ changes sign near $t = 250$ yr. Hence, we expect this time to be a good estimate of the critical time t_m^*. Indeed, when a simulation

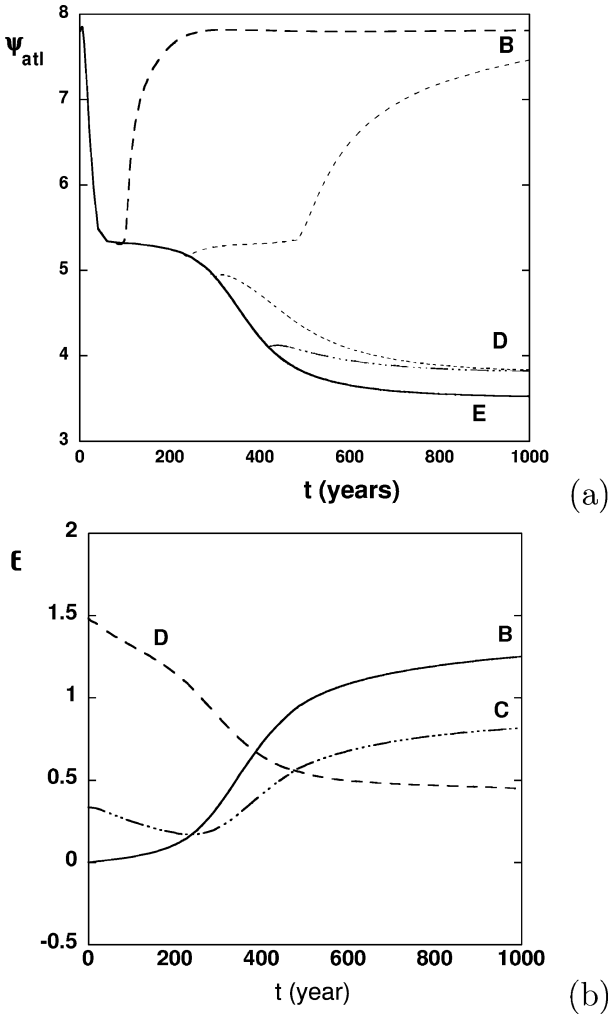

Fig 7. (a) Trajectories of the global model. The dashed curve starts at state B and shows the time evolution under conditions $\gamma_{\mathrm{p}} = -0.23$ for $t_{\mathrm{m}} = \infty$; it ends up for $t \to \infty$ in state E. The solid curve is for $t_{\mathrm{m}} = 100$ and the dash-dotted curve is for $t_{\mathrm{m}} = 400$. (b) Energy anomalies $\mathcal{E}$ of the state along the dashed trajectory in (a) with respect to the states B, C and D.

is performed with $t_{\mathrm{m}} = 100$ yr, the Atlantic circulation recovers after about 100 yr and state B is eventually reached (dashed curve in Fig. 7b). When $t_{\mathrm{m}} = 400$ yr, the Atlantic overturning collapses as seen in the dash-dotted curve in Fig. 7b. The trajectories for the two other times t_{m} (225 and 275 yr) show that $t_{\mathrm{m}}^* \in [225, 275]$. Hence, the estimate of t_{m}^* from the tendency of the energy functional $\mathcal{E}_C$ is excellent.

Note that these results are meant here only to illustrate the methodology and hence we do not put much emphasis on the quantitative results. To obtain useful quantitative estimates of the instability boundaries, the global model should be improved, in particular in its representation of mixing processes.

5. Summary and discussion

Using simple ideas from bifurcation theory, an approach to the TPR problem has been proposed. The central idea is summarized in a more abstract context with help of Fig. 8. In Fig. 8a, two stable fixed points x_1 and x_2 are sketched, which represent stable ocean circulation patterns. In addition, there is an unstable state (which is labeled here x_{s}) that is characteristic for the hysteresis regime in global ocean circulation models (Rahmstorf 1995; Weijer et al. 2003).

In a typical case, the changes in equilibrium states due to a permanent freshwater anomaly, which is represented by a change in parameters of the dynamical system, is as in Fig. 8b. Here, there is only one stable state y, while the states x_1 and x_2 still exist but not as equilibrium states. For example, starting at x_2 the trajectory is attracted to state y after application of the anomalous forcing over a time t_{m} and, depending on t_{m}, it ends up somewhere in phase space, say at z (Fig. 8b). The evolution of the trajectory when the forcing is removed depends on whether this point z is located 'left' (in the attraction basin of x_1) or 'right' (in the attraction basin of x_2) of the state x_{s} of the original bifurcation diagram (Fig. 8a). In the case sketched here, the point z is located 'right' of x_{s} and hence the state x_2 will be reached and no transition occurs. If z would be 'left' of x_{s}, a transition to x_1 occurs.

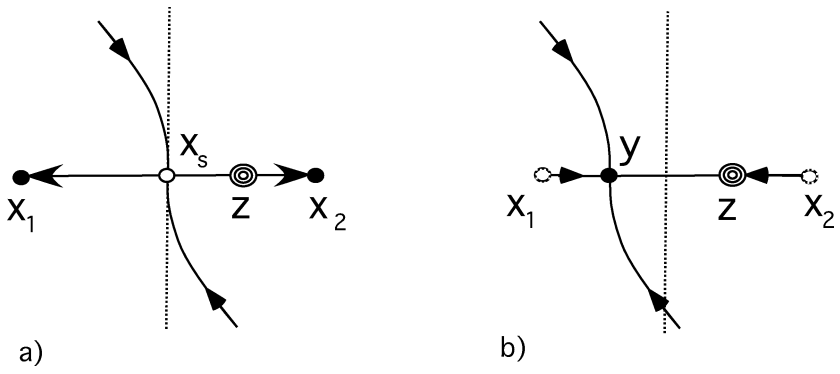

Fig 8. Sketch of the attractors in (a) the original system where x_1 and x_2 are stable fixed points and (b) the dynamical system under a permanent freshwater anomaly where only the state y is a stable fixed point.

While it is, in general, difficult to identify whether the unstable state is crossed at a time $t < t_m$, we have proposed to use energy functionals as indicators. The latter provide a global norm of the difference of the time-dependent solution and the known steady states. For the global model, this provides adequate estimates of the time t_m^* and this indicates that the governing dynamics in phase space is 'near two-dimensional'. As an alternative indicator, one could also take the differences in the Atlantic overturning between the transient solution and the equilibria. The strength of the Atlantic overturning of state C is about 4.8 Sv and this boundary is crossed at about $t = 310$ yr (Figs 6 and 7a).

Crucial for application of this procedure in ocean general circulation models is the computation of the unstable steady states. While for simple box models the unstable states are easily determined, large-scale ocean models which use explicit forward time integration do not have this possibility. With transient flow computations, only stable equilibrium states can be reached within these models. An implicit formulation of ocean models, however, enables an efficient computation of these unstable steady states. This opens the way to use the methodology in this paper to solve the TPR problem in these models and obtain quantitative estimates of the type of freshwater anomalies that may induce a collapse of the thermohaline circulation.

The threshold results are important to determine whether the thermohaline circulation may collapse due to the increase in atmospheric CO_2. Many climate models indicate that an increased greenhouse gas concentration leads to freshwater anomalies on the ocean surface, which subsequently affect the ocean circulation. In Schmittner and Stocker (1999), it is found that with higher emission rates the present thermohaline circulation is more easily destabilized. This result is easily interpreted here: a higher emission rate leads to a faster change in freshwater anomalies and consequently the critical time for reaching the unstable steady state decreases. The results will also be important for the interpretation of relatively rapid climate changes as seen during the last glacial period, such as the Dansgaard–Oeschger oscillations. Also here, changes in the Atlantic overturning circulation due to freshwater perturbations are thought to be crucial (Alley et al. 2003).

6. Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research (NWO) under a PIONIER grant to the first author. All computations were performed on the Origin 3800 computer at SARA, Amsterdam. Use of these computing facilities was sponsored by the National Computing Facilities Foundation (NCF) under the project SC029 with financial support from the NWO.

References

Alley, R. B., Mayewski, P. A., Sowers, T., Stuiver, M., Taylor, K. C. et al. 1997. Holocene climate variability: a prominent widespread event 8200 years ago. *Geology* **25**, 483–486.

Alley, R. B., Marotzke, J., Nordhaus, W. D., Overpeck, J. T., Peteet, D. M., et al. 2003. Abrupt climate change. *Nature* **299**, 2005–2010.

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

Dijkstra, H. A. 2000. *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*. Kluwer Academic, Dordrecht.

Fanning, A. F. and Weaver, A. J. 1997. Temporal-geographical meltwater influences in the North Atlantic conveyor: implications for the Younger Dryas. *Paleoceanography* **12**, 307–320.

Ganopolsky, A., Rahmstorf, S., Petoukhov, V. and Claussen, M. 2001. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158.

Levitus, S., Burgett, R. and Boyer, T. 1994. *World Ocean Atlas 1994, Volume 3: Salinity. NOAA Atlas NESDIS 3*, US Department of Commerce, Washington, DC, pp. 0–99.

Manabe, S. and Stouffer, R. J. 1995. Simulation of abrupt climate change induced by freshwater input into the North Atlantic Ocean. *Nature* **378**, 165–167.

Manabe, S. and Stouffer, R. J. 1999. Are two modes of thermohaline circulation stable?. *Tellus* **51A**, 400–411.

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

Rooth, C. 1982. Hydrology and ocean circulation. *Prog. Oceanogr.* **11**, 131–149.

Schiller, A., Mikolajewicz, U. and Voss, R. 1997. The stability of the North Atlantic thermohaline circulation in a coupled ocean–atmosphere general circulation model. *Climate Dyn.* **13**, 325–347.

Schmittner, A. and Stocker, T. F. 1999. The stability of the thermohaline circulation in global warming experiments. *J. Climate* **12**, 1117–1133.

Schmitz, W. J. 1995. On the interbasin-scale thermohaline circulation. *Rev. Geophys.* **33**, 151–173.

Stocker, T. F., Wright, D. G. and Mysak, L. A. 1992. A zonally averaged, coupled ocean–atmosphere model for paleoclimate studies. *J. Climate* **5**, 773–797.

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

Trenberth, K. E., Olson, J. G. and Large, W. G. 1989. A global ocean wind stress climatology based on ECMWF analyses, Technical Report, National Center for Atmospheric Research, Boulder, CO.

Vellinga, M., Wood, R. A. and Gregory, J. M. 2002. Processes governing the recovery of a perturbed thermohaline circulation in HadCM3. *J. Climate* **15**, 764–780.

Weijer, W. and Dijkstra, H. A. 2001. Bifurcations of the three-dimensional thermohaline circulation: the double hemispheric case. *J. Mar. Res.* **59**, 599–631.

Weijer, W., Dijkstra, H. A., Öksüzoglu, H., Wubs, F. W. and De Niet, A. C. 2003. A fully-implicit model of the global ocean circulation. *J. Comput. Phys.* **192**, 452–470.