Tellus (2002), 54A, 89–98
Printed in UK. All rights reserved

TELLUS
ISSN 0280–6495

On freshwater-dependent bifurcations in box models of the interhemispheric thermohaline circulation

By SVEN TITZ[1,*], TILL KUHLBRODT[1,2], STEFAN RAHMSTORF[2] and ULRIKE FEUDEL[3], [1]*Institute of Physics, Potsdam University, Potsdam, Germany;* [2]*Potsdam Institute for Climate Impact Research, Potsdam, Germany;* [3]*Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Germany*

(Manuscript received 28 March 2001; in final form 10 October 2001)

ABSTRACT

Conceptual box models of the interhemispheric thermohaline circulation are studied with respect to bifurcations. Freshwater fluxes are the main control parameters of the system: they determine the stable states and transitions between stable states of the large-scale thermohaline circulation. In this study of interhemispheric box models both numerical and analytical methods are used to investigate transition mechanisms of the thermohaline circulation. The box model examined first is an interhemispheric four-box model. It is shown that the two bifurcations where the present THC can become unstable, the saddle-node and the Hopf bifurcation, depend in a different way on hemispheric freshwater fluxes. A reduction of the model variables leads to the conclusion that two fixed freshwater fluxes between three surface boxes are the model feature responsible for the bifurcation behaviour found. The significance of the Hopf bifurcation for the stability of the thermohaline circulation is discussed.

1. Introduction

The thermohaline circulation (THC) of the Atlantic ocean (sometimes referred to as the "conveyor belt" is a density-driven large-scale overturning motion with relatively warm surface waters flowing northward and cold North Atlantic deep water returning southward at 2–3 km depth. This circulation carries heat northward at a rate of up to 1 PW (1 PW = 10^{15} W) and has a significant effect on climate, which can be seen e.g. in climate model experiments (Manabe and Stouffer, 1988), or by looking at the winter sea ice margins (Fig. 1 of Rahmstorf, 1997) or the deviations of the climatological air temperature from the zonal mean (Fig. 1 of Rahmstorf and Ganopolski, 1999). The air over the northern North Atlantic is warmed by up to $\sim 10\,^{\circ}$C in annual mean, with the largest effect occurring in winter when oceanic heat release is at its maximum and solar heating at its minimum.

Paleoclimatic reconstructions show that the Atlantic circulation has been subject to large and rapid changes throughout the last Ice Age. Three main circulation modes have been identified in both sediment data and models (Alley et al., 1999; Ganopolski and Rahmstorf, 2001): a warm or interglacial mode with deep water forming in the Nordic Seas and large oceanic heat transport to northern high latitudes (the present climate operates in this mode); a cold or stadial mode with deep water forming south of the shallow sill between Greenland, Iceland and Scotland; and a "switched off" or "Heinrich" mode with practically no deep water formation in the North Atlantic. In the last mode, the Atlantic deep circulation is dominated by inflow of Antarctic bottom water from the south.

* Corresponding author.
e-mail; sven@agnld.uni-potsdam.de

A full hierarchy of ocean and climate models has been used to study the nonlinear behaviour of the Atlantic circulation, its equilibria, stability thresholds and mode transitions (see reviews by Weaver and Hughes, 1992; Rahmstorf et al., 1996 and Rahmstorf, 2000). It was found that the nonlinearity stems mainly from two positive feedbacks: an advective feedback and a convective feedback (Rahmstorf, 1999). Simple box models play an important role in understanding the THC, as they are easy to understand, individual processes and feedbacks can be studied in isolation, and bifurcation maps can often be computed analytically. Qualitative agreement between box models and highly complex circulation models is good in many respects (Rahmstorf, 1996), and box models can be used to interpret results from coupled general circulation models [e.g., the apparent climate instability found by Tziperman (1997) can be reproduced and explained with the help of a box model, Rahmstorf and Ganopolski, 1998]. The present paper is concerned with the nonlinearity of the circulation arising from advective feedback. This feedback was first studied in the seminal box model of Stommel (1961), which consisted of two boxes in one hemisphere. In this model, the stable state of the THC loses its stability at a saddle-node bifurcation (Stommel's bifurcation point). Increasing freshwater forcing (the responsible control parameter) reduces the north–south density difference which determines the overturning rate, while northward salt advection by the overturning circulation counteracts this. At the bifurcation point the northward advection of salty water is no longer able to balance the surface freshwater input to the northernmost box in the model, and the THC breaks down. This basic mechanism occurs in all the variations on Stommel's model which have subsequently been studied (e.g., Rooth, 1982; Marotzke, 1990; Joyce, 1991; Huang and Stommel, 1992; Tziperman et al., 1994).

In addition to Stommel's bifurcation a Hopf bifurcation occurs in some models (Tziperman et al., 1994; Scott et al., 1999), and the THC becomes unstable before the saddle-node bifurcation point is reached.

In this paper we use the box model of Rahmstorf (1996) and modifications of it to investigate systematically the role of freshwater forcing for both saddle-node and Hopf bifurcations. The model has been designed to mimic the interhemispheric THC of the Atlantic ocean. Both numerical and (where possible) analytical bifurcation analyses are performed. We try to make the box model as simple as possible while retaining the key features of its qualitative behaviour [i.e., the topology in phase and parameter space, including bifurcations: see Guckenheimer and Holmes (1983) for an introduction]. Bifurcation points are followed in parameter space and interpreted as instability mechanisms of the box model THC. In section 2 of this paper the basic box model is described briefly, and a numerical bifurcation analysis using path-following software is performed. For changing freshwater fluxes two bifurcations can be found: a Hopf bifurcation and a saddle-node bifurcation. The analytical solutions for these bifurcations are presented for a "minimal" version of the box model in section 3. Section 4 describes the impact of the Hopf bifurcation on the stability of the THC in terms of its basin of attraction in phase space. In the final section the implications of the analysis are discussed.

2. The basic four-box model

2.1. Description

The basic box model we study and modify is Rahmstorf's (1996) interhemispheric four-box model. It has been designed to cover the qualitative behaviour of the large-scale circulation cell of the THC found in general circulation models (GCMs). In Fig. 1 it is shown that two boxes represent the surface and deep water layer in the tropics, whereas one box is set up for the North and South Atlantic, respectively. Mixed boundary conditions are applied, i.e. surface temperatures are relaxed to prescribed values and freshwater

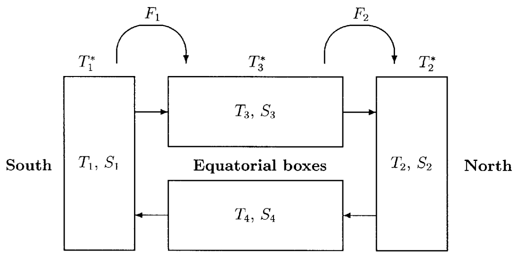

Fig. 1. The basic box model: box 3 represents the tropical surface layer, and box 4 the deep water layer.

fluxes are fixed. The boxes are connected by a flow with volume transport m as indicated by arrows in Fig.1.

For the present circulation direction the model equations read:

$$\dot{T}_1 = \frac{1}{V} m(T_4 - T_1) + \lambda(T_1^* - T_1) \quad (1)$$

$$\dot{T}_2 = \frac{1}{V} m(T_3 - T_2) + \lambda(T_2^* - T_2) \quad (2)$$

$$\dot{T}_3 = \frac{1}{V} m(T_1 - T_3) + \lambda(T_3^* - T_3) \quad (3)$$

$$\dot{T}_4 = \frac{1}{V} m(T_2 - T_4) \quad (4)$$

$$\dot{S}_1 = \frac{1}{V} m(S_4 - S_1) + \frac{1}{V} S_0 F_1 \quad (5)$$

$$\dot{S}_2 = \frac{1}{V} m(S_3 - S_2) - \frac{1}{V} S_0 F_2 \quad (6)$$

$$\dot{S}_3 = \frac{1}{V} m(S_1 - S_3) + \frac{1}{V} S_0 (F_2 - F_1) \quad (7)$$

Every box has a homogeneous temperature T_i and salt content S_i. S_0 is a reference salinity ($S_0 = 35.0$ psu) used to convert the freshwater fluxes into the unit psu s^{-1}.

The salinity of box 4 can be computed from the total salt content S_{tot} and the other salinities because of salt conservation in the model:

$$S_4 = S_{tot} - S_1 - S_2 - S_3 \quad (8)$$

As the dynamical equations (1)–(7) of the model do not depend on absolute salinity values but only on salinity differences, we can use any value for S_{tot}.

The inverse of the temperature restoring coefficient λ is the relaxation time τ, and the T_i^* are the prescribed restoring temperatures. F_1 and F_2 are hemispheric freshwater fluxes which not only represent atmospheric water vapour transport but also wind-driven oceanic transports. The latter is the reason why F_1 in the present climate is a freshwater transport directed towards the Equator (i.e., into the Atlantic; Rahmstorf, 1996; a view which is supported by Weijer et al. 1999), in spite of the Atlantic being an evaporative basin.

The overturning rate m is proportional to the density difference between box 1 and box 2. Density depends linearly on temperature and salinity. Thus, the overturning rate m is

$$m = k[\beta(S_2 - S_1) - \alpha(T_2 - T_1)] \quad (9)$$

where k is a hydraulic constant which is the most tunable parameter. It can be used to fit bifurcation diagrams to simulation runs with global circulation models. Here, we use $k = 23 \times 10^{17}\ \mathrm{m^3\ yr^{-1}}$. This value yields an overturning rate of about 18 Sv (1 Sverdrup $= 1 \times 10^6\ \mathrm{m^3\ s^{-1}}$) when the approximated parameter values of "present climate" (given later) are used. α and β are expansion coefficients for temperature and salinity ($\alpha = 1.7 \times 10^{-4}\ \mathrm{K^{-1}}$ and $\beta = 0.8 \times 10^{-3}\ \mathrm{psu^{-1}}$).

In general, we use the following parameter values: $T_1^* = 0\,°C$, $T_2^* = 3.8\,°C$, $T_3^* = 15\,°C$, $F_1 = 0.05$ Sv (a conservative estimate), $F_2 = 0.25$ Sv and $\tau = 25$ yr, that is $\lambda = 0.04\ \mathrm{yr^{-1}}$. The box volume used for all four boxes is $V = 10^{17}\ \mathrm{m^3}$. Some authors use different volumes for different boxes (Rooth, 1982; Joyce, 1991; Tziperman et al. 1994). For example, they use a smaller box volume for box 2, as the water column of deeply mixed water is less extended compared to the tropical water masses. This can be considered as a more realistic setup. We have also studied the box model with different box volumes but found no difference in the qualitative behaviour. For simplicity we therefore present the results with equal box volumes.

If $(\rho_2 - \rho_1)$ is negative, the advective terms of the model must be adequately reformulated, because the circulation direction is inverse then. In this case, the model equations are

$$\dot{T}_1 = -\frac{1}{V} m(T_3 - T_1) + \lambda(T_1^* - T_1) \quad (10)$$

$$\dot{T}_2 = -\frac{1}{V} m(T_4 - T_2) + \lambda(T_2^* - T_2) \quad (11)$$

$$\dot{T}_3 = -\frac{1}{V} m(T_2 - T_3) + \lambda(T_3^* - T_3) \quad (12)$$

$$\dot{T}_4 = -\frac{1}{V} m(T_1 - T_4) \quad (13)$$

$$\dot{S}_1 = -\frac{1}{V} m(S_3 - S_1) + \frac{1}{V} S_0 F_1 \quad (14)$$

$$\dot{S}_2 = -\frac{1}{V} m(S_4 - S_2) - \frac{1}{V} S_0 F_2 \quad (15)$$

$$\dot{S}_3 = -\frac{1}{V} m(S_2 - S_3) + \frac{1}{V} S_0 (F_2 - F_1) \quad (16)$$

For the given formulation the model equations are not differentiable with respect to m at $m = 0$, but algorithms of numerical bifurcation analysis only work properly with differentiable models. This shortcoming was eliminated by a technical trick: instead of m we use the function

$$m^+ = \frac{m}{1 - e^{-am}} \tag{17}$$

for advection terms with northward surface flow, and the function

$$m^- = \frac{-m}{1 - e^{-am}} \tag{18}$$

for those with southward surface flow. We then use both advection terms in each model equation. The parameter a has no physical meaning. The deviation from the physically correct function m can be made arbitrarily small by increasing the parameter a:

$$\lim_{a\to\infty} m^+ = -\lim_{a\to\infty} m^- = \begin{cases} m & \text{for } m \geq 0 \\ 0 & \text{for } m < 0 \end{cases} \tag{19}$$

Qualitative behaviour, and in particular bifurcation points near $m = 0$, are always checked with respect to the limit $a \to \infty$. In the numerical bifurcation analysis, $a = 10$ is used.

2.2. *Bifurcation study of the basic model*

Conceptual models can contribute to a better understanding of some basic properties of the THC. Although quantitative results cannot be expected to be exact, the occurrence of bifurcations is a rather robust finding from box models. Therefore, a numerical bifurcation analysis of the basic model is performed. We use CANDYS/QA (Feudel and Jansen, 1992) for that purpose.

For this ocean box model, the most important control parameters are the freshwater fluxes F_1 and F_2.

In Fig. 2 the bifurcation behaviour for varied F_1 is displayed. If the southern freshwater flux F_1 is increased, the stable steady state (upper branch with northern sinking) will become unstable at a Hopf bifurcation. It is a subcritical Hopf bifurcation because the emerging cycle is unstable.

The additional bifurcation point shown in Fig. 2 is a saddle-node bifurcation where the stationary state remains unstable. This saddle-node bifurcation is the same bifurcation as in Stommel's box model, where it corresponds to the loss of stability;

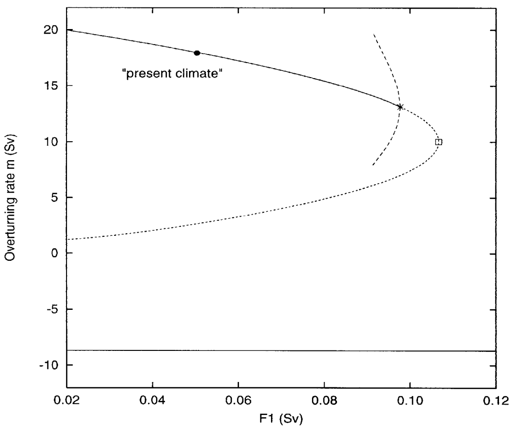

Fig. 2. Upper stable (solid) branch: increasing F_1 leads to a Hopf bifurcation (star), where an unstable periodic solution (dashed) emerges. The points of the unstable periodic solution are the minimum and the maximum overturning rate. At the saddle-node bifurcation (square) instability remains. Lower stable branch: inverse flow of the THC.

the Hopf bifurcation does not exist in Stommel's two-box model.

An unstable periodic solution emerges at the Hopf bifurcation. The advective mechanism which is responsible for the periodic solution is due to the fixed freshwater fluxes. The Hopf bifurcation cannot occur in Stommel's box model, as at least three boxes are needed for the mechanism. Period times of the unstable periodic solution are on a millennial timescale.

If one chooses the values of state variables at the Hopf bifurcation point as initial condition for a simulation with $F_1 > F_{1,\mathrm{Hopf}}$, the new fixed point attractor will be a state with southern sinking and inverse flow of about -9 Sv (lower stable branch). Thus, we have a bistable system for $F_1 < F_{1,\mathrm{Hopf}}$. Decreasing F_1 on the lower stable branch leads to another subcritical Hopf bifurcation (not shown) where the THC switches on again, resulting in a hysteresis behaviour of the model circulation. However, there is a caveat, as this box model with parameters and geometry chosen appropriately for the present climate is unlikely to cover the behaviour of the THC with weak or inverse overturning.

The same bifurcation behaviour also holds with different volumes for different boxes.

In the box model of Scott et al. (1999), Hopf bifurcations also occur on the stable branches with northern and southern sinking. In their bifurcation diagram, the northern freshwater flux is the bifurcation parameter. Scott et al. (1999) show a curve of a transient, unstable solution connecting the two Hopf bifurcations. This transient solution was observed by the authors in critical perturbation experiments. As Scott et al. (1999) state, it is not rigorously defined. Using numerical bifurcation analysis, we find that the emerging unstable cycles are not connected with each other.

The restoring temperature T_2^* can also be used as a control parameter, as the atmospheric temperatures of the northern hemispheric high latitudes will probably increase most in future climate change. The corresponding bifurcation diagram is not shown, because it looks very similar to Fig. 2. For increasing T_2^*, the upper stable branch becomes unstable at a Hopf bifurcation.

With constant temperatures, the box model can be fitted to a perturbation experiment with a global circulation model, as shown in Fig. 7 of Rahmstorf (1996). For this purpose, k and the interhemispheric temperature difference $(T_2 - T_1)$ are tuned.

By following the two bifurcation points of the upper stable branch in two-parameter space of F_1 and F_2, one can study when the Hopf bifurcation occurs. This is shown in Fig. 3. The Hopf bifurcation curve vanishes where it touches the saddle-node bifurcation curve [in a Takens–Bogdanov (TB) point]. It is obvious that the saddle-node bifurcation does not depend on the northern freshwater flux F_2, which can be shown analytically for reduced box models (see section 3). In contrast, the Hopf bifurcation curve is determined by both freshwater fluxes. We can derive a qualitative distinction of paths from present climate (with a stable THC) to instability of the THC with northern sinking from Fig. 3:

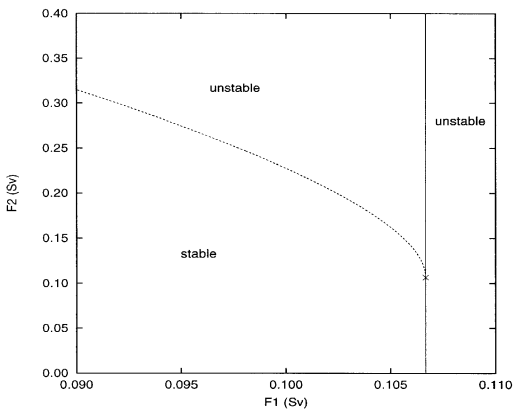

Fig. 3. Two-parameter bifurcation diagram. The Hopf bifurcation point curve (dotted) and the saddle-node bifurcation point curve (solid) meet in a Takens–Bogdanov point (cross).

Increase of F_1 alone; $F_2 < F_{2,TB}$	saddle-node bifurcation
Increase of F_1 alone; $F_2 > F_{2,TB}$	Hopf bifurcation
Increase of F_2 alone	Hopf bifurcation
Decrease of F_1; increase of F_2	Hopf bifurcation
Increase of F_1 and F_2	saddle-node or Hopf bifurcation
Increase of F_1; decrease of F_2	saddle-node or Hopf bifurcation

The outcome of the last two paths depends on the ratio $\dot{F}_1/\dot{F}_2$ and on the initial parameter values.

Following this qualitative picture of a box model, the THC of the present climate (or, generally speaking, the THC in a strong pole-to-pole state) can become unstable due to an increase in one of the hemispheric freshwater fluxes or due to combined changes.

3. Bifurcations in simpler box models

3.1. Description

A four-box model with seven independent variables is in itself a highly conceptual model. Nevertheless, we study simpler modifications in order to find the essential features needed for the bifurcation behaviour of the basic model. For this purpose a model with constant temperatures is considered (as temperature restoring terms are small compared with the advection and freshwater flux terms). In the simplified box model the equatorial deep-water box (box 4) was omitted. Actually, this is very similar to the box model of Rooth (1982), although he used a model with different box volumes for the tropics and the high latitudes. Box 4 can be neglected when steady states and bifurcations are studied, but it seems to be necessary for a better representation of the time-dependent system behaviour.

In the modified model, only two, independent variables are left:

$$V\dot{S}_1 = S_0F_1 + \begin{cases} m(S_2 - S_1) & \text{for } m \geqslant 0 \\ m(S_1 - S_3) & \text{for } m < 0 \end{cases} \tag{20}$$

$$V\dot{S}_2 = -S_0F_2 + \begin{cases} m(S_3 - S_2) & \text{for } m \geqslant 0 \\ m(S_2 - S_1) & \text{for } m < 0 \end{cases} \tag{21}$$

The salinity S_3 is computed from the other salinities:

$$S_3 = S_{\text{tot}} - S_1 - S_2 \tag{22}$$

In the equation for the overturning rate m the temperature difference is now a parameter:

$$m = k[\beta(S_2 - S_1) - \alpha T^*] \tag{23}$$

with $T^* = T_{2,\text{const}} - T_{1,\text{const}}$.

3.2. *Analytical solutions for the bifurcations*

We consider the model with positive overturning rate m which is equivalent to the upper stable branch of the basic model. The stationary state is calculated for the reduced model from eqs. (20) and (21) by solving $\dot{S}_i = 0$, $i = 1, 2$. Then, linear stability theory is applied: the characteristic equation for the eigenvalues of the Jacobian at the stationary state can be solved analytically. It yields equations for the occurrence of local bifurcations of the model (see Appendix).

The qualitative behaviour turns out to be the same as for the basic model. The saddle-node bifurcation is independent of F_2 (as for the basic model):

$$F_1 = \frac{k\alpha^2}{4\beta S_0} T^{*2} \tag{24}$$

The Hopf bifurcation depends on F_1, F_2, and T^*:

$$F_1 = \frac{C}{S_0} + \frac{1}{S_0} \times \left(C^2 + S_0F_2 \frac{3(\alpha^2/\beta)kT^{*2} - S_0F_2}{16} \right)^{1/2} \tag{25}$$

with

$$C = \frac{3}{32}\frac{\alpha^2}{\beta} kT^{*2} - \frac{1}{4} S_0F_2 \tag{26}$$

The curves of the two bifurcation points meet in the TB point which is independent of F_1:

$$T^* = -\left(\frac{2\beta}{k\alpha^2} S_0F_2 \right)^{1/2} \tag{27}$$

Thus, we have an analytic expression for the occurrence of Hopf bifurcations:

$$F_2 > \frac{k\alpha^2}{2\beta S_0} T^{*2} \tag{28}$$

Equation (28) says that the minimum value of the freshwater flux F_2 required for a Hopf bifurcation is proportional to the square of the prescribed interhemispheric temperature difference $T^* = T_{2,\text{const}} - T_{1,\text{const}}$.

The qualitative behaviour of the basic model, i.e. a saddle-node bifurcation and a subcritical Hopf bifurcation which meet in a Takens–Bogdanov point, is fully represented by the simple three-box model with constant temperatures. The model feature which is essential for Hopf bifurcations is the existence of three surface boxes with two connecting freshwater fluxes. Thus, we have a "minimal" interhemispheric box model.

Variable temperatures provide a negative feedback and are important for the quantitative response (Rahmstorf and Ganopolski, 1999 and appendix of Rahmstorf, 1996), but temperatures can be held constant in box model studies of the qualitative behaviour of the THC. This is supported by the fact that the qualitative features of the bifurcation diagrams shown do not change whether the temperatures are variables or not.

4. The unstable cycle of the Hopf bifurcation and the basin boundary

Both the basic box model and the reduced model can exhibit a subcritical Hopf bifurcation. At the bifurcation point an unstable cycle emerges. In the following the role of the unstable cycle is discussed.

The stable state coexisting with the unstable cycle has a certain basin of attraction which can be computed numerically. Every simulation starting with initial conditions within the basin of attraction leads to the stable state of the THC. The unstable cycle turns out to be located on the boundary of this basin. In the reduced model it is the basin boundary itself. This is shown in Fig. 4 for a value of F_1 near the Hopf bifurcation. All initial conditions within the cycle converge to the stable steady state. In a box model with a different

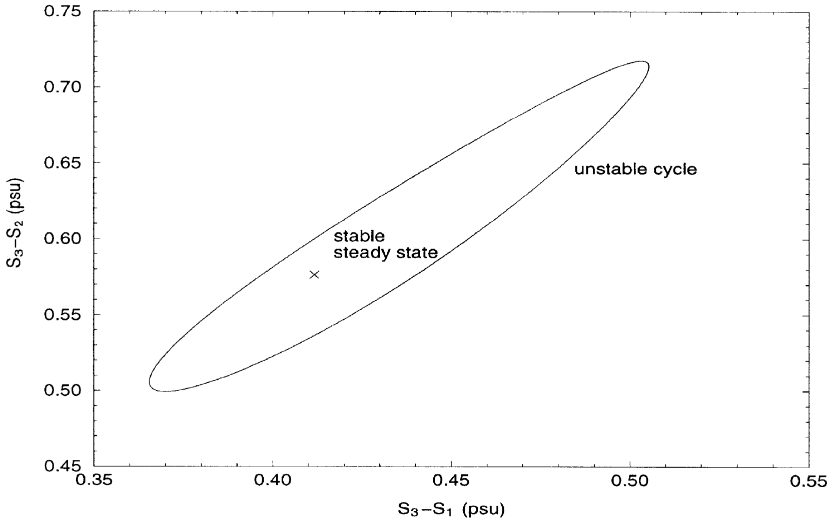

Fig. 4. The unstable cycle is the boundary of the basin of attraction for $F_1 = 0.045$ Sv and $T^* = -2$ K.

heat flux parameterization, Stone and Krasovskiy (1999) investigated this cycle using Van der Pol's method, which yields an equation for the period of the limit cycle.

The period of the unstable periodic orbit strongly depends on the box volumes used. As it is very difficult to define realistic box volumes, we think that periodic behaviour of the THC should not be studied with box models, since the uncertainties are too big to relate model results to observations (e.g. paleodata).

When the control parameter F_1 is increased, the stable state gets closer to the Hopf bifurcation and both the unstable cycle and the basin boundary shrink in size. Thus, there is a *critical radius* of deviations from the stable state: disturbances in the state variables pushing the system beyond that radius would make the THC become unstable *before* F_1 reaches the value of the Hopf bifurcation. In addition, even those disturbances which cause a temporary *increase* of the overturning strength can destabilize the THC and lead to a collapse.

Subcritical Hopf bifurcations do occur in simple box models, but not in 2D fluid models: Quon and Ghil (1995) and Dijkstra and Molemaker (1997) found supercritical Hopf bifurcations on the pole-to-pole branches of their bifurcation diagrams, that is, a *stable* cycle emerges at the bifurcation point. Rivin and Tziperman (1997) studied a coupled box model and found that there are stable oscillations only if the overturning rate is a nonlinear function of the density gradient instead of a linear one. Rahmstorf (1996) found a linear relation between overturning strength and density difference in a GCM study, and this is why we use a linear function here.

Using nonlinear transformations according to normal form theory it can be shown (Titz et al., 2002) that the Hopf bifurcation on the positive overturning branch is always subcritical, i.e., the emerging cycle is always unstable.

5. Conclusions

In this paper, interhemispheric box models of the THC are studied with respect to state transitions when freshwater fluxes are varied. A "minimal" box model of the interhemispheric THC is found. The unstable periodic solution that emerges at one bifurcation point limits the stability of the THC.

Like Stommel's (1961) model, the box models exhibit bistability, i.e., for a certain parameter range both the positive and the negative overturning circulation are stable. The stable state of "present climate" THC (positive overturning) can become unstable due to two bifurcations that are different in nature: a saddle-node bifurcation or a Hopf bifurcation. Scott et al. (1999) also found these bifurcations in their box model. Which of the bifurcations occurs depends on hemispheric freshwater fluxes; this is now shown in a specific bifurcation diagram (Fig. 3). The two bifurcations represent two different mechanisms by which the present THC can become unstable: the saddle-node bifurcation can only occur for a change in total freshwater input into the North Atlantic catchment, but not for a redistribution of freshwater between the low and high latitudes of the Northern Atlantic. The Hopf bifurcation, in contrast, depends on both freshwater fluxes.

We have reduced the variables of the box models to find a "minimal" interhemispheric box model which exhibits both bifurcations described above. At least three boxes with surface contact connected by two hemispheric freshwater flux terms appear to be required for the qualitative behaviour, i.e., the properties of the water masses on a large scale in the Atlantic (two well-mixed high-latitude water columns and the tropical surface layer) are the essential feature needed for the bifurcations we found. Analytical solutions for the two bifurcations are given, so that one can clearly see how they depend on the parameters.

It is not clear a priori which of the bifurcations could lead to a state transition of the THC when freshwater fluxes change in the model. Only beyond a threshold for the freshwater flux in the Northern Hemisphere does the Hopf bifurcation exist in the model. In contrast to Scott et al. (1999) we find that the Southern Hemisphere freshwater flux governs stability in general and that the Northern Hemisphere freshwater flux only plays a role if its value is beyond the threshold mentioned above.

If a Hopf bifurcation is possible, this has consequences for the stability of the THC near the bifurcation point. At the Hopf bifurcation, an unstable periodic orbit emerges which coexists with the stable steady state of the THC with positive overturning. In the "minimal" model, the unstable periodic orbit is identical with the boundary of the basin of attraction that belongs to the stable steady state. As this basin of attraction shrinks when the Hopf bifurcation is approached, small perturbations may destabilize the THC even if the bifurcation point is not yet reached. Positive overturning perturbations can also destabilize the THC in this case.

Therefore, it is not sufficient to study bifurcation points alone: the basin of attraction around stable steady states should also be taken into account when the risk of state transitions is investigated. It is not necessarily at the bifurcation point where state transitions may occur, since small perturbations are always present.

Whether a Hopf bifurcation may possibly cause a destabilization of the THC should be investigated with two- or three-dimensional fluid models. Whether the Hopf bifurcation is super- or subcritical is an important additional question. Up to now, Hopf bifurcations occurring in two-dimensional models were supercritical, i.e., the emerging cycle was stable (Quon and Ghil, 1995; Dijkstra and Molemaker, 1997). Rivin and Tziperman (1997) found with a box model that the Hopf bifurcation is supercritical only if the overturning rate is a nonlinear function of the density gradient, which was not the case in the GCM study of Rahmstorf (1996). In our study we used a linear function, and this is probably why the Hopf bifurcation we find is subcritical. Perhaps a Hopf bifurcation could be found in models of even higher resolution, and especially with strongly asymmetric boundary conditions. A thorough comparison of the underlying mechanism and its dependence on model setup needs to be done in further studies.

Oscillations on a decadal and centennial scale discovered in GCMs and intermediate models are often localized on the North Atlantic (with a decadal time scale), except for the 320 yr oscillation found by Mikolajewicz and Maier-Reimer (1990) in an ocean general circulation model and the 200–300 yr oscillations studied by Mysak et al. (1993) in a two-dimensional ocean model. If a supercritical Hopf bifurcation can be found in a three-dimensional model, with a stable oscillation connected to it, the latter oscillations on a centennial timescale might perhaps be explained in this way.

This study shows that methods of nonlinear dynamics applied to simple ocean box models yield valuable information about different possible transition mechanisms of the THC and their dependence on relevant parameters.

6. Acknowledgements

We thank Wolfgang Jansen for help with the bifurcation analysis software. This work was supported by the Deutsche Forschungsgemeinschaft.

7. Appendix: Bifurcations of the minimal model

After a transformation of variables ($\hat{S}_1 = S_1$, $\hat{S}_2 = S_2 - S_1$), the stationary state of the reduced model is given by

$$\hat{S}_1 = \frac{1}{3}\left(S_{\text{tot}} - \frac{S_0(F_1 + F_2)}{k(\beta\hat{S}_2 - \alpha T^*)}\right) - \hat{S}_2 \quad \text{(A1)}$$

$$\hat{S}_2 = \frac{\alpha T^*}{2\beta} + \left[\left(\frac{\alpha T^*}{2\beta}\right)^2 - \frac{S_0 F_1}{k\beta}\right]^{1/2}. \quad \text{(A2)}$$

By solving $\dot{\hat{S}}_i = 0$ $(i-1, 2)$ the characteristic equation for the eigenvalues of the Jacobian can be computed.

The characteristic equation yields the eigenvalues of the system:

$$\lambda_{1,2} = A \pm \left(A^2 - \frac{3}{V^2}\,km(2\beta\hat{S}_2 - \alpha T^*)\right)^{1/2} \quad \text{(A3)}$$

with

$$A = \frac{1}{2V}\{k\beta[S_{\text{tot}} - 3(\hat{S}_1 + 2\hat{S}_2)] + 3k\alpha T^*\} \quad \text{(A4)}$$

If one real eigenvalue becomes zero, a saddle-node bifurcation exists. Hence, we can find eq. (24) for the saddle-node bifurcation.

At a Hopf bifurcation, two purely imaginary, complex conjugate eigenvalues must exist, i.e., $A = 0$. Thus, we can calculate eq. (25) for Hopf bifurcations.

At a TB point, both conditions (for the saddle-node bifurcation and the Hopf bifurcation) must hold.

REFERENCES

Alley, R. B., Clark, P. U., Keigwin, L. D. and Webb, R. S. 1999. Making sense of millennial scale climate change. In: *Mechanisms of global climate change at millennial time scales* (eds. P. U. Clark, R. S. Webb and L. D. Keigwin), AGU, Washington, 385–394.

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

Dijkstra, H. A. and Molemaker, M. J. 1997. Symmetry breaking and overturning oscillations in thermohaline-driven flows. *J. Fluid Mech.* **331**, 169–198.

Feudel, U. and Jansen, W. 1992. CANDYS/QA — a software system for the qualitative analysis of non-linear dynamical systems. *Int. J. Bifurcation & Chaos* **2**, 773–794.

Ganopolski, A. and Rahmstorf, S. 2001. Simulation of rapid glacial climate changes in a coupled climate model. *Nature* **409**, 153–158.

Guckenheimer, J. and Holmes, P. 1983. *Nonlinear oscillations, dynamical systems and bifurcations of vector fields.* Springer Verlag, Berlin.

Huang, R. X. and Stommel, H. M. 1992. Convective flow patterns in an eight-box cube driven by combined wind stress, thermal and saline forcing. *J. Geophys. Res.* **97**, 2347–2364.

Joyce, T. M. 1991. Thermohaline catastrophe in a simple four-box model. *J. Geophys. Res.* **20**, 393–402.

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

Marotzke, J. 1990. Instabilities and multiple equilibria of the thermohaline circulation. PhD dissertation, Christian-Albrechts-Universität, Kiel, Germany.

Mikolajewicz, U. and Maier-Reimer, E. 1990. Internal secular variability in an ocean general circulation model. *Clim. Dyn.* **4**, 145–156.

Mysak, L. A., Stocker, T. F. and Huang, F. 1993. Century-scale variability in a randomly forced, two-dimensional thermohaline ocean circulation model. *Clim. Dyn.* **8**, 103–116.

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

Rahmstorf, S. 1995. Multiple convection patterns and thermohaline flow in an idealized OGCM. *J. Climate* **8**, 3028–3039.

Rahmstorf, S. 1996. On the freshwater forcing and transport of the Atlantic thermohaline circulation. *Clim. Dyn.* **12**, 1799–1811.

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

Rahmstorf, S. 1999. Rapid transitions of the thermohaline ocean circulation — a modelling perspective. In: *Reconstructing ocean history: a window into the future* (eds. F. Abrantes and A. C. Mix), Kluwer Academic/Plenum Publisher, 139–149.

Rahmstorf, S. 2000. The thermohaline ocean circulation — a system with dangerous thresholds? *Clim. Change* **46**, 247–256.

Rahmstorf, S. and Ganopolski, A. 1998. Simple theoretical model may explain apparent climate instability. *J. Climate* **12**, 1349–1352.

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

Rahmstorf, S., Marotzke, J. and Willebrand, J. 1996. Stability of the thermohaline circulation. In: *The warm water sphere of the North Atlantic ocean* (ed. W. Krauss), Borntraeger, 129–158.

Rivin, I. and Tziperman, E. 1997. Linear versus self-sustained interdecadal thermohaline variability in a coupled box model. *J. Phys. Oceanogr.* **27**, 1216–1232.

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

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

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

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

Titz, S., Kuhlbrodt, T. and Feudel, U. 2002. Homoclinic bifurcation in an ocean circulation box model. *Int. J. Bif. Chaos*, in press.

Tziperman, E. 1997. Inherently unstable climate behaviour due to weak thermohaline ocean circulation. *Nature* **386**, 592–595.

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

Weaver, A. J. and Hughes, T. M. C. 1992. Stability and variability of the thermohaline circulation and its link to climate. In: *Trends in physical oceanography*, Council of Scientific Research Integration, Trivandrum, India, 15–70.

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