On Freshwater-Dependent Bifurcations in Box Models of the Interhemispheric Thermohaline Circulation

Sven Titz, * Till Kuhlbrodt, * Stefan Rahmstorf, ** and Ulrike Feudel *

* *Institute of Physics, Potsdam University, Potsdam, Germany*
** *Potsdam Institute for Climate Impact Research, Potsdam, Germany*

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 breakdown 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 is discussed in terms of the stability of the thermohaline circulation.
An extension of the basic box model, with two circulation cells, is studied with symmetric and asymmetric boundary conditions. The results are compared with those from earlier fluid models. Their bifurcation behaviour, especially the symmetry breaking, can be reproduced in part.

1. Introduction

The thermohaline circulation (THC) of the Atlantic ocean (sometimes referred to as '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 degrees 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 both in sediment data and models (Alley et al. 1999): 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.

A full hierarchy of ocean and climate models has been used to study the non-linear behaviour of the Atlantic circulation, its equilibria, stability thresholds and mode transitions (see reviews by Weaver and Hughes 1992; Rahmstorf, Marotzke, and Willebrand 1996). It was found that the non-linearity 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 thermohaline circulation, 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 non-linearity 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, Marotzke, and Stone 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 systematically investigate 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 analysis is performed. We try to make the box model as simple as possible while retaining the key features of its qualitative behaviour (that is, the topology in phase and parameter space, including bifurcations). 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 section 5 bifurcations of a box model version with two overturning cells are discussed, both for symmetric forcing and after symmetry-breaking. In the final section the implications of the analysis are discussed.

2. The basic 4-box model

a. Description

The basic box model we study and modify is Rahmstorf's (1996) interhemispheric 4-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 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:

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

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

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

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

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

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

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

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 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:

$$m = k\,(\beta\,(S_2 - S_1) - \alpha\,(T_2 - T_1)) \tag{8}$$

where k is a hydraulic constant which is the most tunable parameter; here, we use $k = 23 \times 10^{17}$ m^3 yr^{-1}. This value yields an overturning rate of about 18 Sv (1 Sverdrup $= 1\times10^6$ 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 ($\alpha = 1.7 \times 10^{-4}$ K^{-1} and $\beta = 0.8 \times 10^{-3}$ psu^{-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:

$$S_4 = S_{\mathrm{tot}} - S_1 - S_2 - S_3 \tag{9}$$

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

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 $\lambda = 0.04$ yr^{-1}. The box volume used for all four boxes is $V = 10^{17}$ 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. But we also studied this model with changed box volumes and found no difference in the qualitative behaviour. Therefore, one box volume for all boxes is used here.**

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:

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

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

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

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

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

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

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

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:

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

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.**

b. 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; the Hopf bifurcation does not exist in Stommel's 2-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 time scale.

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 an other subcritical Hopf bifurcation (not shown), where the THC switches on again, resulting in a hysteresis behaviour of the model circulation. But 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 occur on the stable branches with northern and southern sinking, too. 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 is 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 2-parameter space of F_1 and F_2, it can be studied 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 from Fig. 3:

Increase of F_1 alone; $F_2 < F_{2,\mathrm{TB}}$	saddle-node bifurcation
Increase of F_1 alone; $F_2 > F_{2,\mathrm{TB}}$	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 bodels

a. Description

A 4-box-model with 7 independent variables is 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 two different models with constant temperatures are considered (as temperature restoring terms are small compared with the advection and freshwater flux terms). The first one has 4 boxes like the basic model, and in the second one 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.

The model equations of the first modified box model read:

$$V\dot{S}_1 = S_0F_1 + \begin{cases} m\,(S_4 - S_1) & \text{for} \quad m \geq 0 \\ m\,(S_1 - S_3) & \text{for} \quad m < 0 \end{cases} \tag{20}$$

$$V\dot{S}_2 = -S_0F_2 + \begin{cases} m\,(S_3 - S_2) & \text{for} \quad m \geq 0 \\ m\,(S_2 - S_4) & \text{for} \quad m < 0 \end{cases} \tag{21}$$

$$V\dot{S}_3 = S_0\,(F_2 - F_1) + \begin{cases} m\,(S_1 - S_3) & \text{for} \quad m \geq 0 \\ m\,(S_3 - S_2) & \text{for} \quad m < 0 \end{cases} \tag{22}$$

Again, S_4 is computed from the other salinities (equation (9)):

In the second modified model, only 2 independent variables are left:

$$V\dot{S}_1 = S_0F_1 + \begin{cases} m\,(S_2 - S_1) & \text{for} \quad m \geq 0 \\ m\,(S_1 - S_3) & \text{for} \quad m < 0 \end{cases} \tag{23}$$

$$V\dot{S}_2 = -S_0F_2 + \begin{cases} m\,(S_3 - S_2) & \text{for} \quad m \geq 0 \\ m\,(S_2 - S_1) & \text{for} \quad m < 0 \end{cases} \tag{24}$$

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

$$m = k\,(\beta\,(S_2 - S_1) - \alpha T^*), \quad \text{with } T^* = T_{2,\text{const}} - T_{1,\text{const}} \tag{25}$$

b. 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 most reduced model from (23) and (24) 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 models (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) and of the number of boxes:

$$F_1 = \frac{k\alpha^2}{4\beta S_0} T^{*2} \tag{26}$$

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

$$F_1 = \frac{C}{S_0} + \frac{1}{S_0}\sqrt{C^2 + S_0 F_2 \frac{3\,\frac{\alpha^2}{\beta} k T^{*2} - S_0 F_2}{16}} \tag{27}$$

with

$$C = \frac{3}{32}\frac{\alpha^2}{\beta} k T^{*2} - \frac{1}{4} S_0 F_2 \tag{28}$$

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

$$T^* = -\sqrt{\frac{2\beta}{k\alpha^2} S_0 F_2} \tag{29}$$

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

$$F_2 > \frac{k\alpha^2}{2\beta S_0} T^{*2} \tag{30}$$

These equations look very similar for the reduced model with 4 boxes. Equation (30) 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 $T^* = T_{2,\mathrm{const}} - T_{1,\mathrm{const}}$.

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 3-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), 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 two reduced models 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 most 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 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. This disagreement between models of different complexity remains an open question.

5. Box models with two circulation cells

The large scale structure of the present THC appears to be one big circulation cell, but a superposition of two hemispheric circulation cells is conceivable as well (Thual and McWilliams 1992). Hence, **if we want to understand the basic mechanisms of a THC consisting of two circulation cells, we**

need to study the qualitative behaviour of a box model with two overturning cells.

Fig. 5 shows the box model used for this purpose. There are two overturning rates m_n and m_s now, which are proportional to the hemispheric density differences. They can be represented as an asymmetric and a symmetric overturning:

$$m_\mathrm{n} = k_2\left(\beta\left(S_2-\frac{S_3+S_4}{2}\right)-\alpha\left(T_2-\frac{T_3+T_4}{2}\right)\right) \quad (31)$$

$$m_\mathrm{s} = k_2\left(\beta\left(\frac{S_3+S_4}{2}-S_1\right)-\alpha\left(\frac{T_3+T_4}{2}-T_1\right)\right) \quad (32)$$

$$m_\mathrm{asym} = \frac{1}{2}\left(m_\mathrm{s}+m_\mathrm{n}\right) = \frac{1}{2}k_2\left(\beta\left(S_2-S_1\right)-\alpha\left(T_2-T_1\right)\right) \quad (33)$$

$$m_\mathrm{sym} = m_\mathrm{n}-m_\mathrm{asym} = \frac{1}{2}\left(m_\mathrm{s}-m_\mathrm{n}\right) \quad (34)$$

The flow from box 4 to box 3 in the tropics is the difference between the hemispheric overturning rates:

$$\begin{aligned} m_\mathrm{t} &= m_\mathrm{n}-m_\mathrm{s} \\ &= k_2\left(\beta\left(S_1+S_2-S_3-S_4\right)-\alpha\left(T_1+T_2-T_3-T_4\right)\right) \end{aligned} \quad (35)$$

Again, the technical trick of the basic model is used:

$$m_i^+ = \frac{m_i}{1-e^{-am_i}} \qquad (i=\mathrm{s},\mathrm{n},\mathrm{t}) \quad (36)$$

Thus, we can formulate the dynamical equations of the model:

$$\begin{aligned}
V\dot{T}_1 &= m_\mathrm{s}^+\left(T_4-T_1\right)+m_\mathrm{s}^-\left(T_3-T_1\right)+V\lambda\left(T_1^*-T_1\right) && (37)\\
V\dot{T}_2 &= m_\mathrm{n}^+\left(T_3-T_2\right)+m_\mathrm{n}^-\left(T_4-T_2\right)+V\lambda\left(T_2^*-T_2\right) && (38)\\
V\dot{T}_3 &= m_\mathrm{s}^+\left(T_1-T_3\right)+m_\mathrm{n}^-\left(T_2-T_3\right)+m_\mathrm{t}^+\left(T_4-T_3\right)+V\lambda\left(T_3^*-T_3\right) && (39)\\
V\dot{T}_4 &= m_\mathrm{n}^+\left(T_2-T_4\right)+m_\mathrm{s}^-\left(T_1-T_4\right)+m_\mathrm{t}^-\left(T_3-T_4\right) && (40)\\
V\dot{S}_1 &= m_\mathrm{s}^+\left(S_4-S_1\right)+m_\mathrm{s}^-\left(S_3-S_1\right)+S_0F_1 && (41)\\
V\dot{S}_2 &= m_\mathrm{n}^+\left(S_3-S_2\right)+m_\mathrm{n}^-\left(S_4-S_2\right)-S_0F_2 && (42)\\
V\dot{S}_3 &= m_\mathrm{s}^+\left(S_1-S_3\right)+m_\mathrm{n}^-\left(S_2-S_3\right)+m_\mathrm{t}^+\left(S_4-S_3\right)+S_0\left(F_2-F_1\right) && (43)
\end{aligned}$$

In this box model, the hydraulic constant k_2 equals 46×10^{17} m^3 yr^{-1} ($k_2=2\,k$). The hydraulic constant is proportional to the inverse of horizontal distance between the centers of overturning-relevant boxes (Huang et al. 1992), and as m is split up into hemispheric overturning rates m_n and m_s here, the horizontal distance is half the value for the interhemispheric rate m. The meaning of the other parameters is the same as for the basic box model, and conservation of total salt content holds as well.

At first, symmetric conditions are considered, i.e. $F_1 = -F_2$ and $T_1^* = T_2^*$. The bifurcation diagram Fig. 6 shows the stable and unstable states which occur when the strength of the symmetric freshwater fluxes $F = -F_1 = F_2$ is varied. The diagram is very similar to the schematic representation for the stationary states of a fluid model with symmetric boundary conditions (Thual and McWilliams 1992, Fig. 3 (b); see also Quon and Ghil 1992). Thual and McWilliams (1992) used fixed fluxes for the temperatures instead of restoring conditions, and their model includes both viscosity and diffusion. Hence, their model is not fully comparable with the box model described here. Nevertheless, our box model exhibits a similar qualitative behaviour as a 2D fluid model. Thual and McWilliams (1992) also compare the qualitative behaviour of their fluid model with box models but do not show any exact bifurcation diagram with the freshwater strength as control parameter.

For values of F in the center of its interval, two asymmetric and two symmetric (overlapping in Fig. 6) stable circulation states are shown. The asymmetric stable states correspond to Thual and McWilliams's (1992) superposition principle, where one circulation cell is salinity-driven and one is thermally driven.

The multistability of the system can be seen better in Fig. 7 which corresponds to Fig. 3 (a) in Thual and McWilliams (1992). Here, the symmetric overturning rate m_{sym} is plotted against the freshwater flux strength F. Pole-to-pole circulation states overlap in this case.

In Fig. 8, the bifurcation diagram (axes like Fig. 6) after a symmetry breaking in the restoring temperatures is shown. All stable states still exist, although strongly distorted. The bifurcation structure has split up: now, due to an imperfection of the pitchfork bifurcation, two separate branches of stationary states exist. Whereas in the symmetric case four stable states were possible for a broad range of the parameter F (0.02 Sv $< F < 1.5$ Sv), this range shrinks after symmetry breaking (0.02 Sv$< F < 0.38$ Sv). A similar result has been found by Dijkstra and Neelin (2000) in a study with a simple coupled ocean-atmosphere model of the zonally averaged THC. They argue that the two stable states of the THC during paleoclimatic conditions might be the northward pole-to-pole circulation state and the distorted thermally driven circulation state.

6. Conclusions

Interhemispheric box models with one and with two circulation cells of the THC were studied with respect to their bifurcations when freshwater fluxes are varied.

Box models of the first type exhibit bistability. The stable state of "present climate" THC can become unstable due to a saddle-node bifurcation or due to a Hopf bifurcation. Which of these occurs depends on hemispheric freshwater fluxes; this is shown in a specific bifurcation diagram. **The two bifurcations represent two different mechanisms how the present THC can become unstable: the saddle-node bifurcation can only occur for a changed freshwater input into the Atlantic, but not for a different freshwater flux between the tropical and the Northern Atlantic. The**

Hopf bifurcation depends on both freshwater fluxes.

We have reduced the variables of box models of the first type to find a "minimal" interhemispheric box model which exhibits both bifurcations described above. Three boxes with surface contact and two hemispheric freshwater fluxes appear to be responsible for the qualitative behaviour. **Analytical solutions for the two bifurcations are given, so that one can clearly see how they depend on the parameters.**

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. 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.

Whether a Hopf bifurcation may possibly cause a destabilization of the THC should be investigated with 2D or 3D fluid models. Up to now, Hopf bifurcations occurring in 2D models were supercritical, that is, the emerging cycle was stable (Quon and Ghil 1995; Dijstra and Molemaker 1997). Oscillations on a decadal and centennial scale found in GCMs and intermediate models often are localized on the North Atlantic (with a decadal time scale) except 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, Stocker and Huang (1993) in a 2D ocean model. It is possible that a Hopf bifurcation can be found in models of higher resolution in a different parameter regime, 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.

A box model of the second type (2 circulation cells) is studied with symmetric boundary conditions and after symmetry breaking. Close similarity to the bifurcation behaviour of Thual and McWilliams' (1992) fluid model can be shown for symmetric boundary conditions. The range of multistability is **strongly** reduced by symmetry breaking which **is surprisingly consistent** with a study using a zonally averaged coupled ocean-atmosphere model (Dijkstra and Neelin 2000).

This study shows that methods of Nonlinear Dynamics applied to simple ocean box models yield valuable information about different breakdown mechanisms of the THC and their dependence on relevant parameters.

Acknowledgements.

The authors wish to thank Wolfgang Jansen for help with the bifurcation analysis software. This work was supported by the Deutsche Forschungsgemeinschaft.

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 most reduced model is:

$$\hat{S}_1 = \frac{1}{3}\left(S_{\mathrm{tot}} - \frac{S_0\,(F_1 + F_2)}{k(\beta\hat{S}_2 - \alpha T^*)}\right) - \hat{S}_2 \tag{44}$$

$$\hat{S}_2 = \frac{\alpha T^*}{2\beta} + \sqrt{\left(\frac{\alpha T^*}{2\beta}\right)^2 - \frac{S_0 F_1}{k\beta}} \tag{45}$$

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:

$$\lambda_{1,2} = A \pm \sqrt{A^2 - \frac{3}{V^2}\,km\,(2\beta\hat{S}_2 - \alpha T^*)} \qquad \text{with} \tag{46}$$

$$A = \frac{1}{2V}\,(k\beta\,(S_{\mathrm{tot}} - 3(\hat{S}_1 + 2\hat{S}_2)) + 3k\alpha T^*) \tag{47}$$

If one real eigenvalue becomes zero, a saddle-node bifurcation exists. Hence, we can find the equation (26) for the saddle-node bifurcation.

At a Hopf bifurcation, two purely imaginary, complex conjugate eigenvalues must exist, that is, $A = 0$. Thus, we can calculate the equation for Hopf bifurcations (27).

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

REFERENCES

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

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

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

Dijkstra, H. A., and J. D. Neelin, 2000: Imperfections of the thermohaline circulation: Latitudinal asymmetry and preferred northern sinking. *J. Climate*, 13, 366-382.

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

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

Huang, R. X., J. R. Luyten, and H. M. Stommel, 1992: Multiple equilibrium states in combined thermal and saline circulation. *J. Phys. Oceanogr.*, 22, 231-246.

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

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

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

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

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

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

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

Quon, C., and M. Ghil, 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. *Reconstructing ocean history: A window into the future*, F. Abrantes, and A. C. Mix, Eds., 139-149.

Rahmstorf, S., and A. Ganopolski, 1998: Simple theoretical model may ex-

plain apparent climate instability. *J. Climate*, 12, 1349-1352.
Rahmstorf, S., and A. Ganopolski, 1999: Long-term global warming scenarios computed with an efficient coupled climate model. *Clim. Change*, 43, 353-367.
Rahmstorf, S., J. Marotzke and J. Willebrand, 1996: Stability of the thermohaline circulation. *The warm water sphere of the North Atlantic ocean*, W. Krauss, Ed., Borntraeger, Stuttgart, 129-158.
Rooth, C., 1982: Hydrology and ocean circulation. *Prog. Oceanogr.*, 11, 131-149.
Scott, J. R.,J. Marotzke, and P. H. Stone, 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 Y. P. Krasovskiy, 1999: Stability of the interhemispheric thermohaline circulation in a coupled box model. *Dyn. Atmos. Oceans*, 29, 415-435.
Thual, O., and J. C. McWilliams, 1992: The catastrophe structure of thermohaline convection in a two- dimensional fluid model and a comparison with low-order box models. *Geophys. Astrophys. Fluid Dyn.*, 64, 67-95.
Tziperman, E., 1997: Inherently unstable climate behaviour due to weak thermohaline ocean circulation. *Nature*, 386, 592-595.
Tziperman, E., J. R. Toggweiler, Y. Feliks, and K. Bryan, 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 T. M. C. Hughes, 1992: Stability and variability of the thermohaline circulation and its link to climate. *Trends in physical oceanography*, Council of Scientific Research Integration, Trivandrum, India, 15-70.
Weijer, W., W. P. M. de Ruijter, H. A. Dijkstra, and P. J. van Leeuwen, 1999: Impact of interbasin exchange on the Atlantic overturning circulation, *J. Phys. Oceanogr.*, 29, 2266-2284.

Figure Captions

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

Figure 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.

Figure 3: 2-parameter bifurcation diagram. The Hopf bifurcation point curve (dotted) and the saddle-node bifurcation point curve (solid) meet in a Takens-Bogdanov point (cross).

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

Figure 5: Box model with two circulation cells.

Figure 6: Symmetric and asymmetric circulation states are shown. The restoring temperatures at high latitudes are equal: $T_1^* = T_2^* = 1°$C. Stable states are plotted as solid lines, unstable states as dotted lines. Filled squares are pitchfork bifurcations, and empty squares are saddle-node bifurcations. On the line $m_{\text{asym}} = 0$ Sv, two stationary states overlap between the two saddle-node bifurcations.

Figure 7: The multiple equilibria of the THC are shown in this diagram of m_{sym} versus F. The asymmetric pole-to-pole circulation states overlap. Empty squares are saddle-node, filled squares pitchfork bifurcations.

Figure 8: Box model with 2 circulation cells after symmetry breaking in the restoring temperatures: $T_1^* = 3°$C, $T_2^* = 1°$C. Empty squares are saddle-node-bifurcations. The upper and the lower branch do not meet at ($m_{\text{asym}} = 0$ Sv, $F = 0$ Sv).

Note to the technical editor:
All *figures are available in electronic form.*

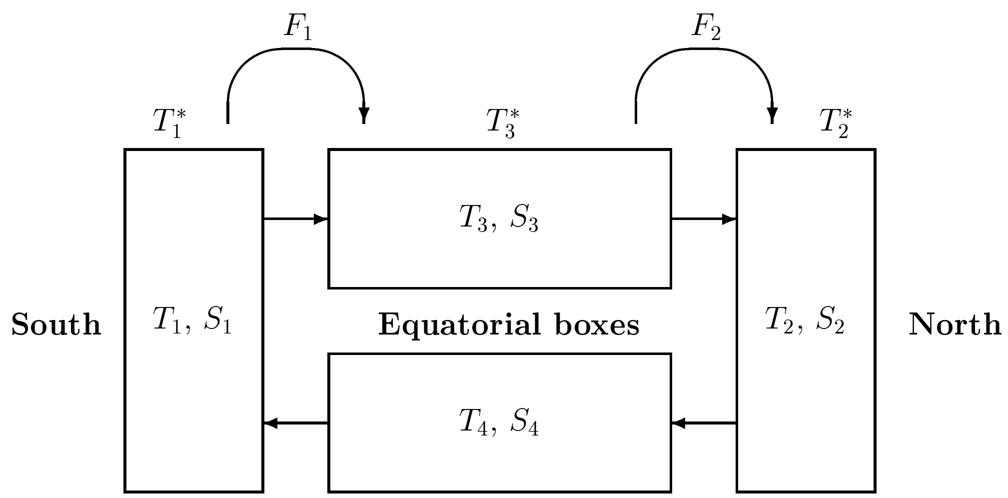
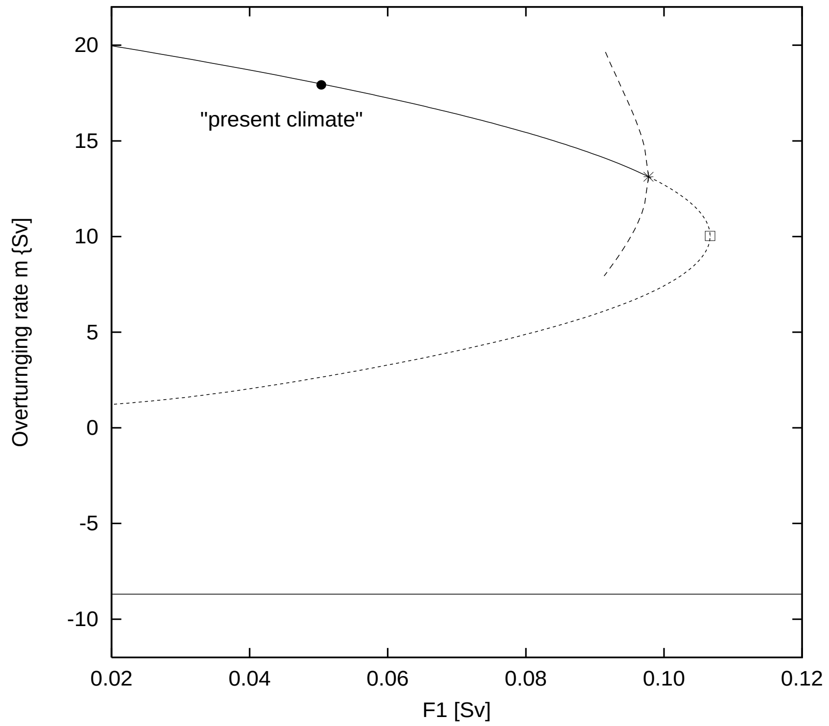


Figure 1: The basic box model: Box 3 represents the tropical surface layer, and box 4 the deep water layer.



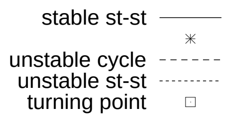


Figure 2: Upper stable (solid) branch: Increasing F_1 leads to a Hopf bifurcation (star), where an unstable periodic solution (dashed) emerges. The points of the unstable periodic solution are the minimum and the maximum overturning rate. At the saddle-node bifurcation (square) instability remains. Lower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

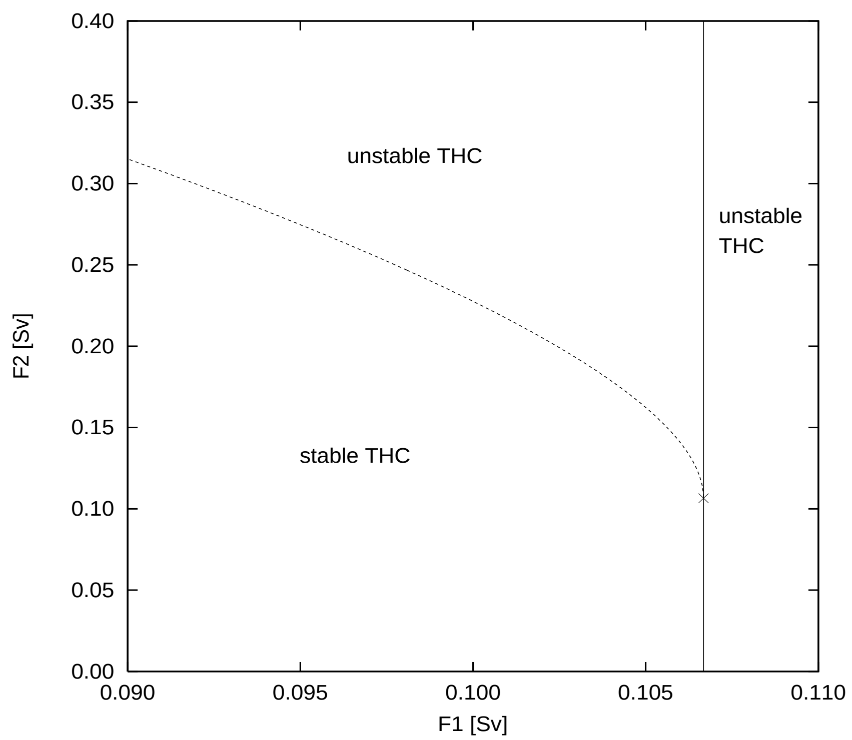


Figure 3: 2-parameter bifurcation diagram. The Hopf bifurcation point curve (dotted) and the saddle-node bifurcation point curve (solid) meet in a Takens-Bogdanov point (cross).

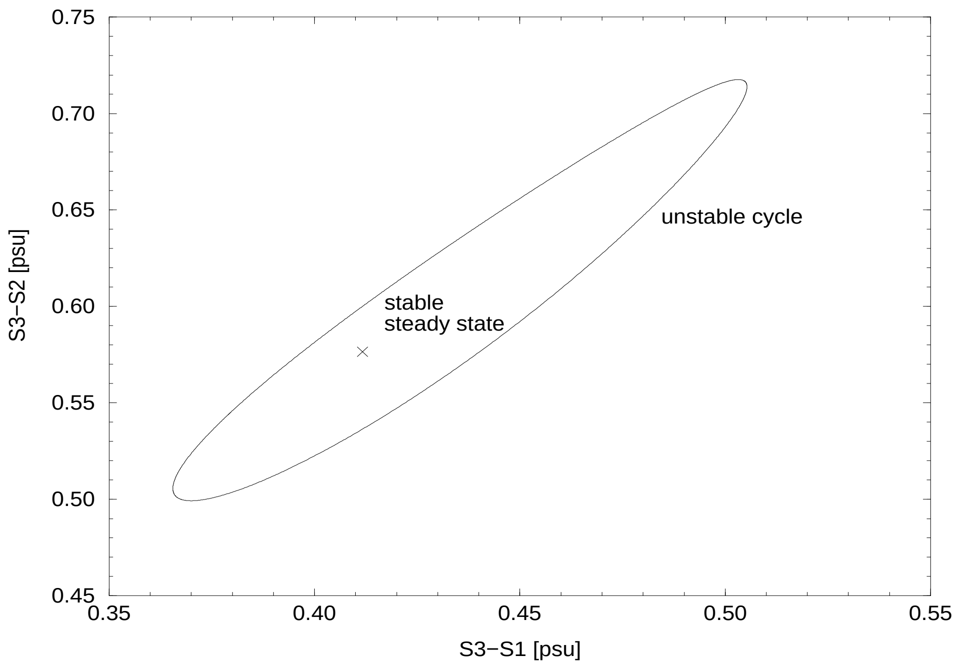


Figure 4: The unstable cycle is the boundary of the basin of attraction for $F_1 = 0.045$ Sv, and $T^* = -2$ K.

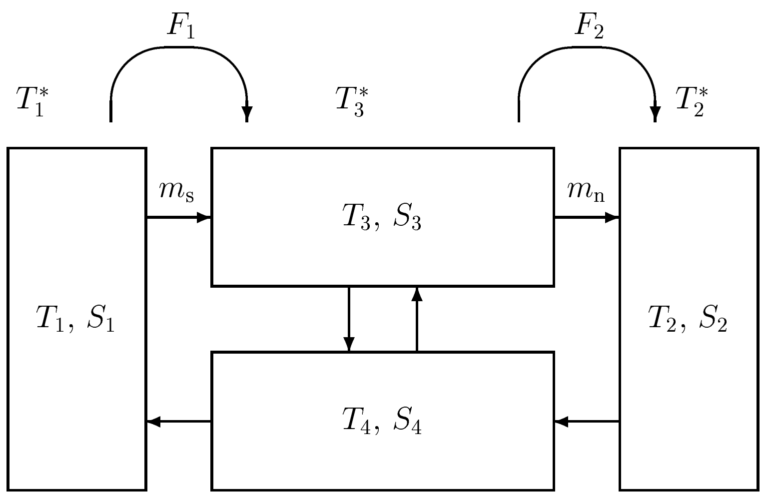


Figure 5: Box model with two circulation cells.

Figure 6: Symmetric and asymmetric circulation states are shown. The restoring temperatures at high latitudes are equal: $T_1^* = T_2^* = 1°\mathrm{C}$. Stable states are plotted as solid lines, unstable states as dotted lines. Filled squares are pitchfork bifurcations, and empty squares are saddle-node bifurcations. On the line $m_{\mathrm{asym}} = 0$ Sv, two stationary states overlap between the two saddle-node bifurcations.

Figure 7: The multiple equilibria of the THC are shown in this diagram of m_{sym} versus F. The asymmetric pole-to-pole circulation states overlap. Empty squares are saddle-node, filled squares pitchfork bifurcations.

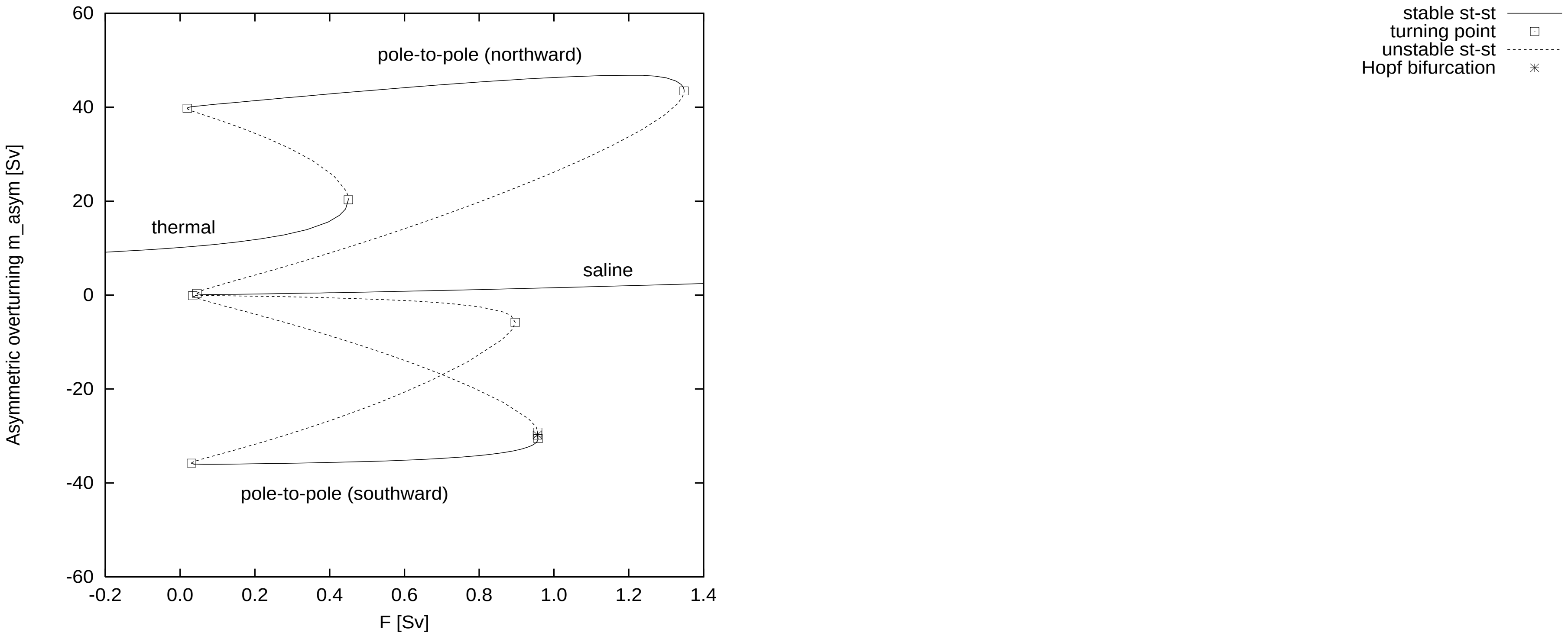


Figure 8: Box model with 2 circulation cells after symmetry breaking in the restoring temperatures: $T_1^* = 3°\mathrm{C}$, $T_2^* = 1°\mathrm{C}$. Empty squares are saddle-node-bifurcations. The upper and the lower branch do not meet at ($m_{\mathrm{asym}} = 0$ Sv, $F = 0$ Sv).