

Thomas Kleinen · Hermann Held
Gerhard Petschel-Held

The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation

Received: 6 June 2002 / Accepted: 8 October 2002
© Springer-Verlag 2003

Abstract A simple two-box model of the hemispheric thermohaline circulation (THC) is considered. The model parameterizes fluctuations in the freshwater forcing by a stochastic process. The dependence of the power spectral density and the lifetime of quasistationary states of the THC on the distance to the bifurcation point, where the THC collapses, is calculated analytically. It is shown that power spectral properties change as the system is moved closer to the bifurcation point. These changes allow an estimate of the distance to the bifurcation point.

Keywords Thermohaline circulation · Stochastic climate model · Power spectrum · Indicator · Bifurcation

1 Introduction

It is becoming increasingly evident that there are critical thresholds in the Earth system, where the climate may change dramatically (Scheffer et al. 2001; Smith et al. 2001). The exact positions of these thresholds are, however, still unclear and it might be doubted whether they can be determined with enough precision to give concrete information on the threat of crossing the threshold. Therefore, additional independent methods for assessing the closeness of the system to these thresholds are needed. These methods could contribute to an early warning system for assessing the danger of crossing a threshold and possibly provide the information necessary for controlling the system.

One subsystem that displays such a threshold is the North Atlantic thermohaline circulation (THC) (Clark

et al. 2002). Considering the system of the THC under a global warming scenario, the circulation may collapse if certain threshold values in northern Atlantic temperature and salinity are exceeded. This system has been thoroughly investigated using a whole range of models, ranging from conceptual models (Stommel 1961; Cessi 1994; Rahmstorf 1996; Scott et al. 1999; Titz et al. 2002) over models of intermediate complexity (Stocker and Schmittner 1997; Rahmstorf and Ganopolski 1999a) to highly complex general circulation models (GCM) (Manabe and Stouffer 1988; Rahmstorf 1995; Schiller et al. 1997). While most of these investigations agree with respect to the fact that there is a threshold where the circulation breaks down, the exact value of this threshold in the climate system has not yet been determined. Currently, it seems questionable whether the exact position of the threshold will be determined within the near future, as the disagreement on overturning strength and sensitivity to freshwater fluxes is still quite large between different GCMs. In addition, it has not been possible to accurately measure the present overturning strength of the THC in the real climate system so far.

Simple box models have shown a remarkable ability to capture important aspects of the behavior of the THC manifested in GCM experiments. So far, most investigations have concentrated either on the deterministic behavior of the THC or on the stochastic properties. Deterministic models have mainly been used to investigate the bifurcations and attractors of the models (Stommel 1961; Rahmstorf 1996; Scott et al. 1999; Titz et al. 2002), whereas the work with stochastic models has concentrated on the spectrum and on the stationary distribution (Stommel and Young 1993; Cessi 1994; Bryan and Hansen 1995). Combining the two points of view, Timmermann, Lohmann, and Monahan have investigated how the stationary distribution changes as a function of the bifurcation parameter (Timmermann and Lohmann 2000; Monahan 2002; Monahan et al. 2002). What has not been investigated so far is the

Responsible Editor: Donna L. Witter

T. Kleinen (✉) · H. Held · G. Petschel-Held
Potsdam-Institute for Climate Impact Research (PIK),
PO Box 601203, 14412 Potsdam, Germany
e-mail: kleinen@pik-potsdam.de

dependence of the power spectrum on the bifurcation parameter.

The concept of stochastic climate models goes back to a paper by Hasselmann (Hasselmann 1976). He observed that there are many fast processes (e.g. weather) within the climate system. These processes may affect the long-term development of the system, so that they cannot be omitted from an assessment. On the other hand, these processes cannot be incorporated into comprehensive models due to resolution, computation, and conceptual restraints. Therefore, Hasselmann proposed that the influence of the fast processes on climate could be modeled as a stochastic forcing to the system.

With respect to bifurcations, the influence of fluctuations becomes even more important, as fluctuations may increase in the vicinity of bifurcation points, eventually leading to critical fluctuations (Haken 1980; Sornette 2000). This might induce a switch in the system even before reaching the bifurcation itself (Monahan 2002).

Within the THC system, the freshwater flux serves as the bifurcation parameter. The freshwater flux is composed of a multitude of components, including precipitation and wind-driven transports among others. These processes are subject to short-term processes and fluctuations. We therefore might have the situation of critical fluctuations.

The aim of this investigation is to see whether the system's response to fluctuations can be used as an indicator for the proximity of the system to the bifurcation point, and to determine whether this might yield additional information compared to purely deterministic methods. Potential applications of this method could span a large range, ranging from the THC itself over other climatic processes to ecosystem dynamics.

The structure of the paper is as follows: in Section 2 we will describe the model used in this investigation and estimate the variability of North Atlantic freshwater fluxes. In Section 3 we will investigate the dependence of the power spectral density on the bifurcation parameter. Section 4 will concentrate on the dependence of the probability density on the bifurcation parameter. In Section 5 we will analyze the stability of the present-day THC and its dependence on bifurcation parameter and noise strength, while we will try to extend our findings to more comprehensive models in Section 6. The paper will finish with summary and conclusions in Section 7.

2 Model description and behavior

2.1 The Stommel model

The Stommel model (Stommel 1961) is a well-known nonlinear conceptual model of the thermohaline circulation. It is a hemispheric two-box model consisting of interconnected boxes of temperature T_i and salinity S_i representing the North Atlantic at low and high latitudes. The model calculates the circulation strength in a channel connecting the two boxes, as well as the changes

in salinity and temperature resulting through advective transport between the boxes. The circulation is proportional to the density difference between the boxes.

In the case of a positive freshwater flux μ , corresponding to an excess of precipitation over evaporation in the high northern latitudes, the Stommel model has two equilibrium states. First, there is a state with strong positive overturning, where the flow is driven by thermal gradients and braked by haline gradients. In addition, there is a second state with weak negative overturning, where the flow is driven by haline gradients.

If the system is in the thermally driven state with positive overturning, an increase in the mean freshwater flux μ leads to a decrease in circulation strength, until a critical freshwater flux μ_c is reached, where the circulation breaks down in a saddle-node bifurcation and the system switches to the haline-driven equilibrium solution. A subsequent reduction of the freshwater flux μ does not result in an instantaneous return to the positive overturning solution, but the system shows hysteresis behavior and stays on the haline branch at first (Monahan 2002).

2.2 Model reduction

We modify the model by adding a stochastic freshwater flux to it. In order to allow an analytical treatment, we reduce the system to one dimension. We will leave out the details of the model formulation and the model reduction, but refer the reader to Monahan (2000), as we are using a model formulation that is nearly identical to his. The only difference is that his model also contains a red-noise term affecting the advective transport which we left out.

If we assume that the temperature relaxation time scale τ is very short, temperature relaxation occurs instantly, constraining the nondimensional temperature gradient x to $x = 1$ if one considers the long time scales only. Under this condition, the behavior of the system on short time scales may be of minor importance for the long-term evolution of the slow variables. In this case, variables can be separated, "slaving" the fast variables to the slow variables. By this 'adiabatic elimination', we reduce the dimensionality of the system to one. Independent of the adiabatic reduction, the qualitative properties of the system, namely the saddle-node bifurcation and the hysteresis behavior, are retained in the reduced system. We therefore consider the one-dimensional stochastic differential equation (SDE) in the (nondimensional) salinity gradient y :

$$dy = (-|1 - y|y + \mu) dt + \sigma dW . \quad (1)$$

In this equation y is the nondimensional salinity gradient between the boxes, μ is the nondimensional freshwater flux, $(1 - y)$ is the (nondimensional) overturning, and σdW is a Wiener process (Gardiner 1994) with amplitude σ . The advective salinity transport $-|1 - y|y$ balances the freshwater flux μ . In the nondimensional system, one nondimensional time unit corresponds to

1.66a (Monahan 2002). Therefore, all plots showing times or frequencies can essentially be read in years.

In the case of positive overturning [$(1 - y) > 0 \Leftrightarrow y < 1$], Eq. (1) transforms to

$$dy = (y^2 - y + \mu)dt + \sigma dW, \quad (2)$$

which has the deterministic steady-state solutions

$$y_{0,1} = \frac{1}{2} \mp \sqrt{\frac{1}{4} - \mu}, \quad (3)$$

with y_1 an unstable and $y_0 = 1/2 - \sqrt{1/4 - \mu}$ a stable fixed point.

Please note that the system described by Eq. (1) is very similar to the one investigated in Cessi (1994). The only difference is that Cessi parameterized the volume transport by the square of the density difference and not by the absolute value used here.

This SDE retains the relevant features of the Stommel model. The system undergoes a saddle-node bifurcation at the critical freshwater flux $\mu_c = 0.25$ and observes a hysteresis behavior.

In the following model simulations, the model equations are integrated using an explicit Euler scheme (Kloeden and Platen 1999) with a time step of 0.1 (nondimensional) “years”. Initial experiments have shown that the convergence properties do not improve using smaller time steps.

2.3 Freshwater flux variability

The freshwater flux μ is composed of a multitude of factors. It contains all processes that influence the salinity balance both in the equatorial Atlantic and in the deepwater formation areas in the high northern latitudes, with the exception of freshwater transport by the THC. For the Labrador Sea, which is one of two deepwater formation areas for the North Atlantic THC (Warren 1981), Houghton and Visbeck estimate that the major contributors to the freshwater balance of that area are advective transport, sea-ice melting, continental runoff, precipitation, and evaporation (Houghton and Visbeck 2002). In addition to these, there is also wind-driven transport, which could have an influence. The freshwater flux μ is composed of all of these factors, minus the advective transport by the THC. This transport is described explicitly by the Stommel model.

In order to assess the stochastic amplitude σ , the variance of the quantities composing μ has to be estimated. Estimates for these factors are difficult to obtain and notoriously unreliable, and the variance of these quantities is rarely even considered. Nonetheless, a few estimates exist.

Walsh and Portis estimate from reanalysis data that the standard deviation of annual averages of precipitation P and evaporation E over the North Atlantic is typically about 10–20% of the mean (Walsh and Portis 1999), while Houghton and Visbeck (2002) report that interannual variations in the Labrador current are about

30% of the mean. In addition, there is an estimate of the variability of the winter sea-ice concentration in the Labrador Sea by Deser et al. (2002).

Apparently sea-ice anomalies are preceded by freshwater anomalies (Houghton and Visbeck 2002) and are therefore correlated with these. It also seems plausible that the interannual variations of P and E are not entirely uncorrelated. If precipitation is anomalously large one year, evaporation must also be anomalously large, though not necessarily in the same location. With regard to the advective transport we simply do not know how representative the Labrador current is for the advective freshwater transport within the entire North Atlantic Basin.

The variance of such a sum of random processes is the sum of the variances, if the processes are independent and uncorrelated. If the processes are correlated, on the other hand, the variance is much harder to quantify.

Therefore, it is only possible to give a minimum estimate for the stochastic amplitude σ that drives the ocean model by using the variability of precipitation P as an indicator for the total variability, as the sea-ice concentration and advective transport influences and their correlations are very difficult to quantify.

Precipitation is usually assumed not to be uncorrelated white noise, but correlated red noise with an autocorrelation e-folding time or decorrelation time scale of a couple of days. In the model (Eq. 1) we have substituted a white noise process for the red noise process of precipitation. This is possible if the time scales of the processes differ widely.

The noise amplitude σ of an equivalent white noise process driving the ocean model can be assessed by comparing the variances of two simple linear models: one that is driven by white noise and one that is driven by red noise. The model driven by white noise σW has variance $\text{var}(y) = \sigma^2/2\delta$ with δ the inverse decorrelation time scale of the model, whereas a model driven by a red noise process z with variance $\text{var}(z) = \lambda^2/2\alpha$ has the variance $\text{var}(y') = \lambda^2/[2\alpha\delta(\alpha + \delta)]$ with α the inverse decorrelation time scale of the red noise process z driving the model, δ the same as for the model driven by white noise and λ the amplitude of the white noise process generating the red noise. From comparing these quantities, it becomes obvious that $\sigma^2 = \lambda^2/[\alpha(\alpha + \delta)]$ and therefore

$$\sigma = \sqrt{\frac{2}{(\alpha + \delta)} \text{var}(z)}.$$

Walsh and Portis report the standard deviation of annual averages of precipitation. The variance of such annual averages is, of course, much lower than the variance of the original process. If we assume the original process to be a piecewise constant time series, we can estimate that the variance of a process \bar{z} , where N values have been averaged, is $1/N$ times the variance of the original process z . Numerical experiments show that this relationship holds well for the actual red noise

process, if a segment length of twice the decorrelation time scale τ is assumed. This gives the relation $\text{var}(z) = N\text{var}(\bar{z})$ for the variances, with N given by $N = (T/2\tau) = \alpha T/2$ and T the averaging time.

Walsh and Portis estimate that the standard deviation of precipitation and evaporation is about 10–20% of the mean, corresponding to a variance $\text{var}(\bar{z}) = (\beta\mu)^2$ with μ the mean freshwater flux and $0.1 \leq \beta \leq 0.2$. If we assume that rain has a decorrelation time scale τ of about a week, $\tau = 1/\alpha \approx 1/50$ yr and therefore $\alpha \approx 50$ yr⁻¹. The inverse decorrelation time scale δ of the ocean model can be assessed from the linearized model presented in Section 3. From Eq. (5) it is obvious that $\delta = 2\sqrt{\Delta\mu}$. As $0 \leq \Delta\mu \leq 0.25$, we obtain $0 \leq \delta \leq 1$. From these considerations it follows that

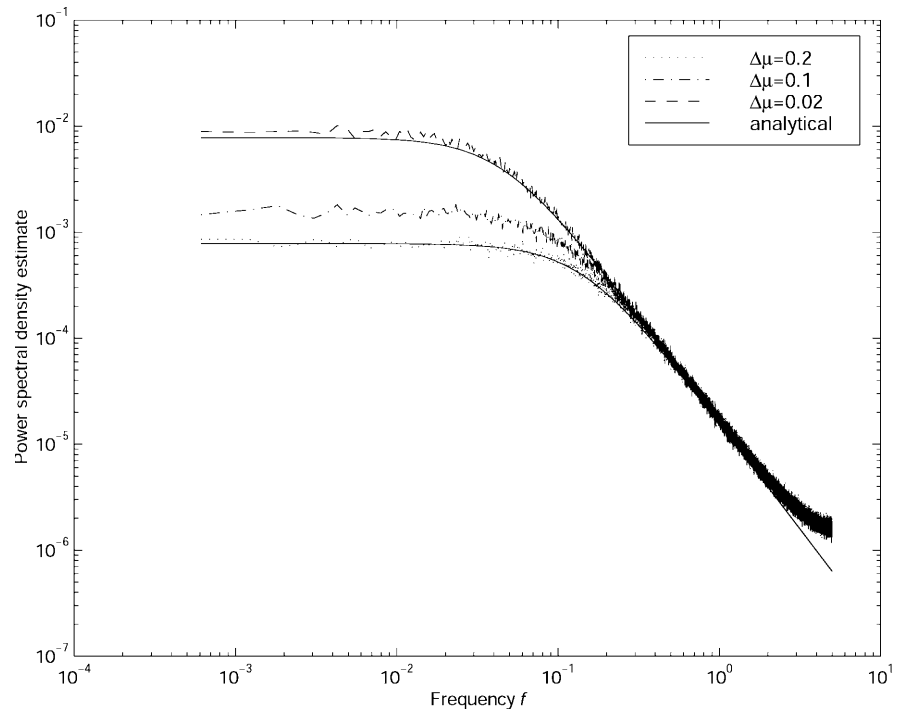
$$\sigma = \sqrt{\frac{2}{(\alpha + \delta)} \frac{\alpha}{2} T (\beta\mu)^2} \approx \sqrt{T} \beta \mu . \quad (4)$$

The maximum mean freshwater flux μ in the reduced Stommel model is $\mu = \mu_c = 0.25$, giving a range from $\sigma = 2.5 \times 10^{-2}$ to $\sigma = 5 \times 10^{-2}$ for the white noise amplitude σ . We are therefore using a value of $\sigma = 2.5 \times 10^{-2}$ in all model simulations, unless another value is specified explicitly.

3 Spectral changes at the bifurcation

In the deterministic case, it is possible to assess the distance to the bifurcation point on the basis of a measured steady-state overturning q . This would require, however, a perfect knowledge of q itself, as well as of all the model parameters.

Fig. 1 Spectral density estimate of overturning in adiabatically reduced model at distances to bifurcation point $\Delta\mu = 0.2$, $\Delta\mu = 0.1$, $\Delta\mu = 0.02$, and corresponding theoretical spectrum. The frequency denoted on the *abscissa* refers to nondimensional time



The stochastic model formulation, on the other hand, reveals additional information because fluctuations are included, the properties of which might be used for indicating the distance to the bifurcation.

The spectrum of the system described by Eq. (2) can be calculated analytically by using small noise expansion (Gardiner 1994), essentially a linearization of the system around the steady-state solution $y_0 = \langle y(t) \rangle$. The prerequisite for applying this approximation is that the noise amplitude is sufficiently small, i.e. the influence of the noise on the system behavior is small compared to the deterministic influences.

The dynamics of the perturbation $\tilde{y} = y - y_0$ is given by the linearization of Eq. (2) around y_0 :

$$\dot{\tilde{y}} = (2y_0 - 1)\tilde{y} + \sigma\tilde{W} . \quad (5)$$

Equation (5) describes an Ornstein–Uhlenbeck Process. The spectrum of this process is (Gardiner 1994):

$$S(\omega) = \frac{\sigma^2}{(2y_0 - 1)^2 + \omega^2} . \quad (6)$$

Using the deterministic steady-state solution (Eq. 3) and the distance $\Delta\mu = \mu_c - \mu$ from the critical freshwater flux μ_c , we obtain the dependence of the power spectral density on the distance to the bifurcation point $\Delta\mu$:

$$S(\omega, \Delta\mu) = \frac{\sigma^2}{4\Delta\mu + \omega^2} . \quad (7)$$

Plotting the power spectral density calculated from the time series of overturning data, a shift in spectral properties due to the lurking bifurcation can be seen. It is clearly visible from Fig. 1 that the spectrum changes as the system is moved towards the bifurcation point.

Far away from the bifurcation, one obtains a red spectrum with a cutoff frequency (the frequency where the spectrum changes from a horizontal to a decreasing shape) of approximately 10^{-1} (please note that frequency is given in nondimensional units, as time is nondimensional), while the magnitude of the spectrum in the limit $\omega \rightarrow 0$ is approx. 8×10^{-4} . Close to the bifurcation, the cutoff frequency decreases to approximately 2×10^{-2} at $\Delta\mu = 0.02$, while the magnitude of the spectrum in the limit $\omega \rightarrow 0$ increases by about an order of magnitude to approximately 8×10^{-3} .

Obviously, the decorrelation time of the overturning increases as one gets closer to the bifurcation, which results in a change in cutoff frequency. This also implies that the probability density function widens and that the amplitude of fluctuations increases.

In Fig. 1 both measured and theoretically estimated spectra of the overturning are shown for different distances to the bifurcation point $\Delta\mu$. Comparing these, it is obvious that both measured and calculated spectrum agree very closely.

By transforming the system to an Ornstein–Uhlenbeck process, we are now capable of discussing the mechanism leading to increased decorrelation times.

Physically, a larger mean freshwater flux μ , corresponding to a smaller distance to the bifurcation point $\Delta\mu$, leads to an increase of the steady-state salinity gradient y_0 (see Eq. 3 for comparison). The increase in y_0 , in turn, reduces the overturning $q = 1 - y$. Thus, the salt-advection feedback (Rahmstorf et al. 1996) that stabilizes the circulation is decreased in strength and small deviations from the steady state change the advection term in Eq. (1) very little, which leads to a very slow relaxation to the steady state, thus allowing larger deviations from y_0 and increasing the decorrelation time. The power spectral density changes towards a “redder” spectrum. The magnitude of the spectrum in the limit $\omega \rightarrow 0$ is inversely proportional to $\Delta\mu$, while the cutoff frequency is proportional to the square root.

4 Probability density at the bifurcation

In order to gain further insight into the processes that take place as the system approaches the bifurcation, it seems valuable to also look into the changes of the probability density function (PDF) of the process. While the Langevin equation describes the temporal evolution of the system itself as a diffusion process, the Fokker–Planck equation describes the temporal evolution of the PDF. The Fokker–Planck equation is equivalent to the Langevin SDE, but it examines a different aspect of the system properties.

The Fokker–Planck equation,

$$\frac{\partial}{\partial t} p(y, t) = \hat{L}_{\text{FP}} p(y, t) \quad (8)$$

is a partial differential equation describing the temporal evolution of the probability density function p . \hat{L}_{FP} is the Fokker–Planck operator

$$\hat{L}_{\text{FP}} = -\frac{\partial}{\partial y} [-\nabla f(y)] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial y^2} \quad (9)$$

with

$$f(y) = -\left(\frac{y^3}{3} - \frac{y^2}{2} + \mu y\right) \quad (10)$$

being the potential for the adiabatically reduced system (Eq. 2) in the vicinity of the potential minimum. If the system is close to the steady-state solution y_0 , a steady-state probability density function can be calculated from the Fokker–Planck equation using the potential solution (Gardiner 1994). This solution is

$$p_{\text{stat}}(y) = N \exp\left[-\frac{2}{\sigma^2} f(y)\right] \quad (11)$$

with a normalization constant N and the potential (10). An analogue to this system is the motion of an overdamped particle in a potential well driven by a Brownian forcing.

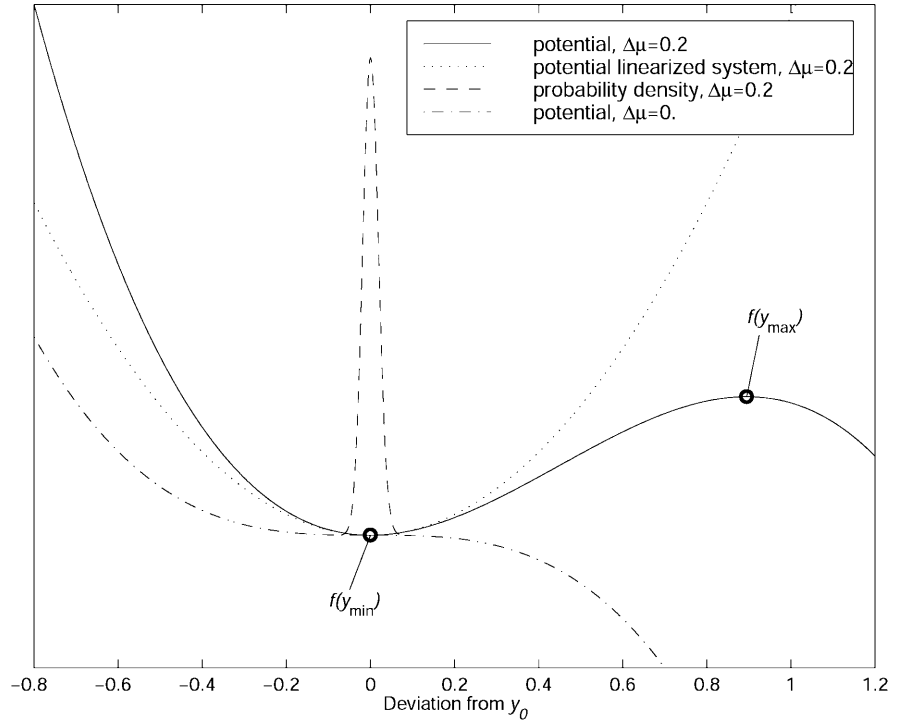
We assume that we can use this *Ansatz* as long as the mean first exit time from the quasistationary state is much longer than the time scales under consideration. Strictly speaking, this method of solution is not applicable because the system is not in a stationary state, but only in a quasistationary state. As outlined in Section 5, this assumption is valid as long as the system is not so close to the bifurcation that a breakdown of the circulation is inevitable.

The adiabatically reduced system (Eq. 2) corresponds to the cubic potential (Eq. 10). Far away from the bifurcation point, there is a well-defined potential well. This is shown quite clearly in Fig. 2, where the potential is plotted for $\Delta\mu = 0.2$ and $\Delta\mu = 0$. At $\Delta\mu = 0.2$ the PDF is tightly constrained within the potential well. As the system moves closer to the bifurcation, the potential difference $\Delta f = f(y_{\text{max}}) - f(y_{\text{min}})$ decreases, until it vanishes at the deterministic bifurcation point μ_c . The PDF broadens and the probability current across the local maximum $f(y_{\text{max}})$ increases.

The approach to a deterministic saddle-node bifurcation corresponds to a decrease in the potential difference between local minimum (corresponding to the stable fixed point in the deterministic system) and local maximum (corresponding to the unstable fixed point) of the stochastic potential, until the potential difference is zero at the deterministic bifurcation point. This flattening of the potential again reflects the vanishing restoring force.

The linearized system (Eq. 5) corresponds to a parabolic potential. As the system approaches the bifurcation, the parabola widens. Therefore, the change in the spectrum can be explained by the widening of the parabolic potential, whereas the bifurcation itself is the vanishing of the potential well.

Fig. 2 Potential and probability density function at distance to bifurcation point $\Delta\mu = 0.2$ and potential at the bifurcation point $\Delta\mu = 0$. The potentials have been transformed by an additive constant, so that they agree in y_0 . Noise amplitude σ was $\sigma = 2.5 \times 10^{-2}$. Potential minimum $f(y_{\min})$ and potential maximum $f(y_{\max})$ are marked with circles. The PDF is obviously constrained to an area where potential and linearization still agree



5 Stability of the THC

As the stochastic system approaches the bifurcation, its stability decreases gradually. Even if the deterministic bifurcation point has not been reached, there still is a finite probability that the stochastic system will leave the potential well and cross over to the haline-driven reverse equilibrium circulation (Monahan 2002).

If the goal of detecting a lurking bifurcation is to provide means for control, the bifurcation has to be detected at a distance large enough from the bifurcation point. Very close to the bifurcation, however, the system will not be stable for very long. If this method is to make sense, the circulation has to be stable long enough to provide enough time for a meaningful measurement and for moving the system away from the bifurcation again, if it is too close to it. Therefore, it is necessary to assess the dependence of the life time of the quasistationary state of the THC on the bifurcation parameter. In addition, the stationary solution developed in the last section does not make any sense if the quasistationary system is not stable on the time scales under consideration.

Within the framework of Fokker–Planck equation and probability density, the life time of the quasistationary state of the system can be described by the mean first exit time from the potential well. In this section we investigate how the mean first exit time depends on the bifurcation parameter.

We consider some arbitrary area in configuration space that contains the current maximum of the PDF. The time when the system leaves this area for the first time in the ensemble mean is the mean first exit time.

The mean first exit time from the potential well can be calculated from the probability density p and the probability current S . The probability current can also be understood as the probability divided by the mean first exit time τ . Therefore, the mean first exit time can be calculated from

$$\tau = \frac{p}{S} = \frac{2}{\sigma^2} \int_{y_1}^{y_2} \exp\left[-\frac{2}{\sigma^2} f(y)\right] dy \int_{y_{\min}}^A \exp\left[\frac{2}{\sigma^2} f(y)\right] dy, \quad (12)$$

with $f(y)$ given by Eq. (10) (Risken 1996). The probability that the system is within the potential well is determined by the first integral in Eq. (12), whereas the inverse of the probability current from the potential minimum at y_{\min} to a point A outside the potential well is estimated in the second integral. The integration boundaries $y_{1,2}$ are the boundaries of the area in configuration space considered, while A is located outside the potential well, somewhat larger than y_{\max} . The integrals in Eq. (12) can be evaluated numerically.

If the potential difference $\Delta f = f(y_{\max}) - f(y_{\min})$ is much larger than the diffusion coefficient $D = \sigma^2/2$, it is also possible to obtain an analytical expression for the mean first exit time by using Kramers' formula (Risken 1996). Thus, we can gain an insight into the dependence of the system's stability on the distance from the bifurcation point. In our case, Kramers' escape time is

$$\tau(\sigma, \Delta\mu) = 2\pi(4\Delta\mu)^{-\frac{1}{2}} \exp\left[\frac{1}{3\sigma^2}(4\Delta\mu)^{\frac{3}{2}}\right]. \quad (13)$$

The mean first exit time τ grows more strongly than exponentially with distance $\Delta\mu$, while it decreases with

rising noise variance σ^2 . As one can see from Fig. 3, this approximation is not valid very close to the bifurcation point. In that case, the mean first exit time has to be calculated numerically.

In order to compare these results with model simulations, we have run an ensemble of 1000 model simulations for every parameter combination considered. The model Eq. (2) was integrated forward in time, and the time when the system left the potential well was considered the first exit time. The mean first exit time τ is given by the ensemble mean. These simulations were performed at noise amplitudes of $\sigma = 2.5 \times 10^{-2}$, $\sigma = 5 \times 10^{-2}$ and $\sigma = 7.5 \times 10^{-2}$ with mean freshwater fluxes μ ranging from $\mu = 0.23$ to $\mu = 0.25$, corresponding to $\Delta\mu = 0.02$ and $\Delta\mu = 0$.

The results obtained by numerical integration of Eq. (12) and by model simulation were nearly identical. Thus, we have plotted the mean first exit times obtained by numerical integration and by the analytical approximation in Fig. 3. From this figure it is obvious that the analytical approximation is not applicable below $\Delta\mu = 1.1 \times 10^2$ at $\sigma = 7.5 \times 10^{-2}$, while it is applicable longer if the noise strength is lower. At larger distances from the bifurcation point all methods give similar answers.

When it comes to a discussion of the stability of the THC, three time scales are of major interest. First, a mean first exit time of 10^4 years or larger is relevant, as that is the age of the Holocene, and the THC has been relatively stable for that long, as can be inferred from North Atlantic sediment data (Bond et al. 1997). The distance to the bifurcation point corresponding to this value of the mean first exit time is the minimum distance

possible for preindustrial conditions. The second relevant time scale is the time needed to obtain a meaningful measurement, while the third time scale is the time needed to change the mean freshwater input μ , if it is determined that we are too close to a breakdown of the THC. We estimate both the second and the third time scale to be in the range of 10^2 – 10^3 years.

As an illustration of the information that can be gained by using this *Ansatz*, we will estimate the distances to the bifurcation point that correspond to these time scales. The analysis is, of course, not directly applicable to the real THC, but only to our simplified model. For convenience, we now assume that a nondimensional time step is equal to 1 year. The error introduced by this simplification is small, as the actual length of a time step is equal to 1.66 yr.

From Fig. 4, where the mean first exit time obtained analytically is shown depending on the distance to the bifurcation point $\Delta\mu$ and on the noise amplitude σ , we can see that a mean first exit time of 10^4 years is reached at $\Delta\mu = 0.07$, corresponding to $\mu = 0.18$, if the noise standard deviation is comparatively high at $\sigma = 7.5 \times 10^{-2}$. With $\sigma = 5 \times 10^{-2}$, this time is reached at $\Delta\mu \approx 0.033$, corresponding to $\mu = 0.217$, while $\Delta\mu \approx 0.013$ at $\sigma = 2.5 \times 10^{-2}$.

A mean first exit time of 10^3 years is reached at $\Delta\mu = 0.052$ for $\sigma = 7.5 \times 10^{-2}$. Therefore, the mean freshwater flux μ must be smaller than $\mu = 0.198$, if we want to be able to still measure and control the system. This minimum permissible distance from the bifurcation point decreases with smaller noise standard deviations to $\Delta\mu = 0.023$ at $\sigma = 5 \times 10^{-2}$ and $\Delta\mu = 0.008$ for $\sigma = 2.5 \times 10^{-2}$.

Fig. 3 Nondimensional mean first exit times in the vicinity of the bifurcation point, calculated analytically and numerically for three noise amplitudes σ

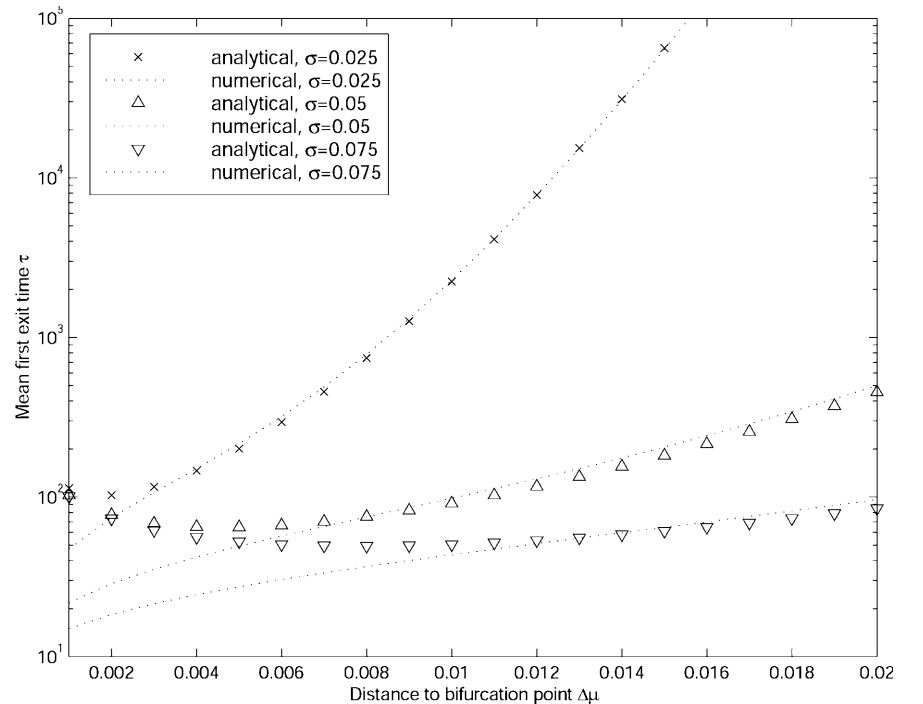
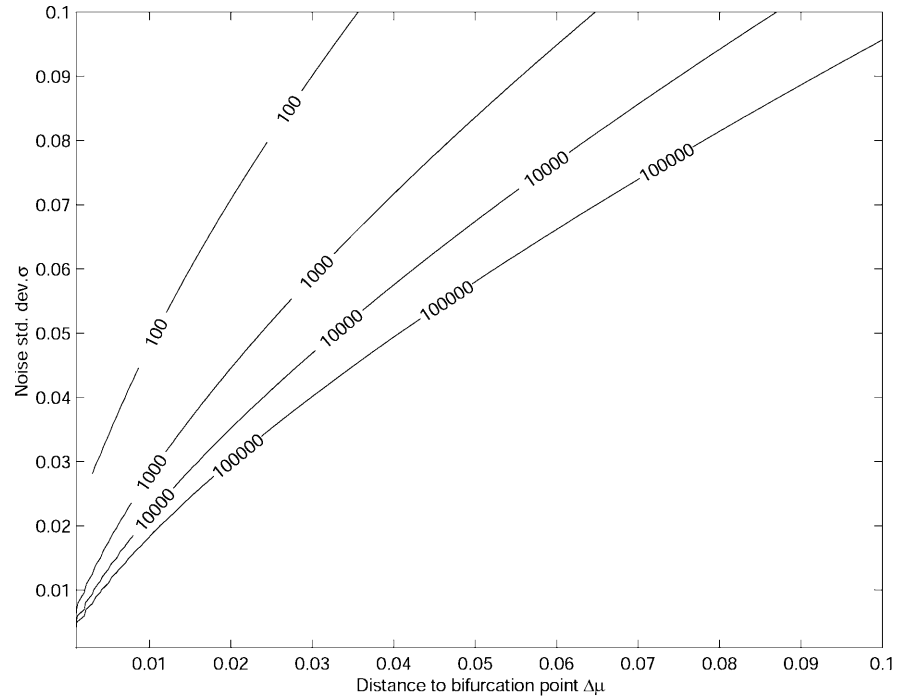


Fig. 4 Nondimensional mean first exit time from the potential well for varying noise strength σ and distance to bifurcation point $\Delta\mu$. Calculated from the analytical approximation where it is valid. The mean first exit time is given in nondimensional time units. Large decorrelation times also allow for the notion of quasistationary states. For sensible parameter combinations, the mean first exit times are always much larger than decorrelation times



6 Beyond the simplified Stommel model

The change in the spectrum as the system is moved closer to the saddle-node bifurcation is caused by a decrease in the strength of the advective feedback stabilizing the THC. The advective feedback mechanism is clearly not a property that is restricted to the simple box model with which we have performed our experiments. Therefore, it can be assumed that the changes in the spectrum we have observed in a very simple box model also occur in more comprehensive models, if the strength of the advective feedback is reduced.

We have performed experiments similar to the ones we described in Section 3 with two less simplified box models. First, we have used the full Stommel model including the temperature dynamics. For the exact model formulation we refer the reader to Monahan (2002) again. The upper panel of Fig. 5 shows the power spectral density at different distances to the bifurcation point $\Delta\mu$. Clearly, the power spectrum changes in a manner similar to the simplified Stommel model as the system gets closer to the bifurcation point.

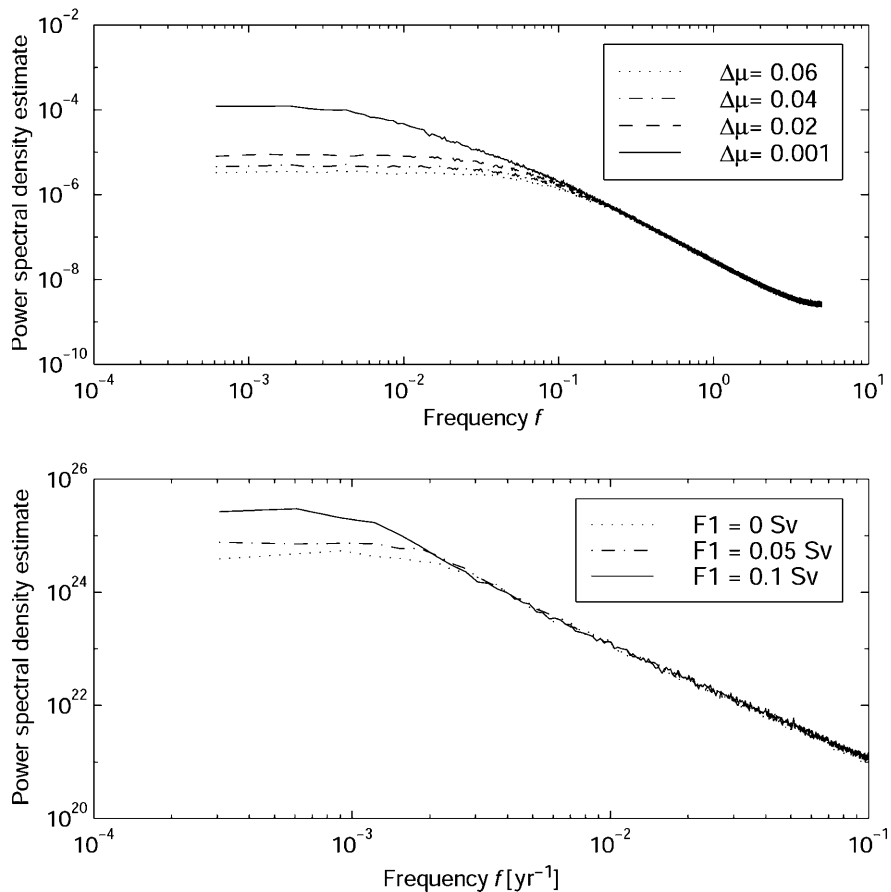
In addition, we have performed these experiments with a salinity-only version of an interhemispheric four-box model of the THC (Titz et al. 2002), which is very similar to the other well-known interhemispheric three- and four-box models (Rooth 1982; Rahmstorf 1996; Rahmstorf and Ganopolski 1999b; Scott et al. 1999; Titz et al. 2002). These models describe an interhemispheric thermohaline circulation, where the strength of the overturning is dependent on the density gradient between the northern and the southern box of the model, while the equatorial box, which is much warmer than the polar boxes, has no

influence on the steady-state circulation, only on the dynamics of changes in circulation strength. In this type of model a modified bifurcation behavior occurs: if there is also a freshwater transport between the equatorial and the northern box of the model that is above a certain threshold, a Hopf bifurcation can occur (Titz et al. 2002). We have modified the model by adding a stochastic freshwater flux between the southern and the equatorial box of the model, and we have performed experiments at different mean freshwater fluxes, while we kept the freshwater flux between the equatorial and the northern box below the critical value for the Hopf bifurcation. The power spectral density of the overturning is shown in the lower part of Fig. 5. Again, there are distinct changes in the power spectrum as the system approaches the bifurcation point.

Due to the lack of a suitable GCM, we have not yet been able to test whether the effect we are describing also occurs in full-scale GCMs, but GCM experiments have shown that the overturning decreases with an increase of the freshwater flux into the Atlantic Basin (Rahmstorf 1996). Decreased overturning implies decreased advection and therefore a decrease in the strength of the advective feedback.

Von Storch et al. have investigated the spectral characteristics of the deep-ocean mass transport (von Storch et al. 2000). They found no evidence against the assumption that the THC has a red spectrum. In fact, they found a power spectral density $S(\omega)$ proportional to ω^{-2} in a frequency range from $\omega = 1/20 \text{ a}^{-1}$ to $\omega = 1/500 \text{ a}^{-1}$. Longer time scales, where a red spectrum would level off, could not be resolved due to the short time series they had available.

Fig. 5 Power spectral density of overturning at different mean freshwater fluxes in less simplified box models. *Upper panel* Full Stommel model, *legend* gives distance to bifurcation point $\Delta\mu$, frequency is nondimensional; *lower panel* four-box model of the interhemispheric THC, salinity only, *legend* gives mean freshwater flux F_1



Knutti and Stocker have recently published results from experiments with a model of intermediate complexity (Knutti and Stocker 2002), where they report that the variability of the overturning increases strongly in the vicinity of the critical threshold where the THC collapses. Tziperman has reported a similar increase in variance close to the instability threshold in experiments with a comprehensive 3-D ocean–atmosphere model (Tziperman 2000). This increase in variability is also predicted by our model.

While all of these characteristics are no conclusive evidence that the spectral behavior we are describing also occurs in GCMs and possibly in the real ocean, they are consistent with the predictions by our model. We therefore feel confident that our results can also be applied to GCMs and the real ocean.

7 Summary and conclusions

In this paper we have demonstrated how the spectral properties and the probability density change as the THC moves closer to the bifurcation point in a stochastic modification of the Stommel model. As the system approaches the bifurcation point, the spectrum becomes “redder”. The magnitude of the power spectral density in the limit of zero frequency increases inversely proportionally to the distance to the bifurcation point

$\Delta\mu$, while the cutoff frequency decreases proportionally to the square root.

The mechanism described above is a generic property of the advective feedback mechanism and the saddle-node bifurcation, and is therefore independent of the exact values of model parameters. The spectral characteristics of the overturning in GCMs and the observed increase in variance of the overturning close to the instability threshold support the hypothesis that the mechanism described above is also valid for more comprehensive models, even though this would have to be confirmed by experiments. The method presented above could have a wide range of applicability to all systems that contain a saddle-node bifurcation or a feedback mechanism similar to the advective feedback of the THC.

Rahmstorf estimated the distance to the bifurcation point by fitting a deterministic four-box model to a GCM hysteresis run of the THC (Rahmstorf 1996). A similar approach could also be used to estimate the distance to the bifurcation point in the real climate system, but it needs the total overturning strength as an input, which is very difficult to measure accurately. Any error in measuring the total overturning would lead to an error in the estimate of the distance to the bifurcation point, which makes an additional independent approach desirable.

Rather than average values of total overturning, our method requires the cutoff frequency of red overturning

power spectra as an input, or, equivalently, the decorrelation time in overturning time series. Therefore, the proposed method allows an independent estimate of the actual overturning, for a twofold reason.

First, the cutoff frequency only requires decorrelation properties of overturning time series as an input, not the total overturning, nor its total variability. Therefore, a time series of a representative fraction of overturning would be sufficient. In addition, fluctuations of salinity and temperature differences show the same change in the power spectral density as one finds close to the bifurcation. Therefore, the distance to the bifurcation point could also be estimated by measuring the fluctuations in either of these quantities.

Second, the proposed method adds information with respect to another source of uncertainty: results obtained with a simple model have to be transferred back to the real world system. This transfer is problematical with both methods, the one proposed here, and the “traditional” based on time-averaged, total overturning estimates. The model parameters, which contain considerable uncertainty, facilitate this transfer. From a simple argument of units it becomes evident that our proposed method is sensitive to an alternative set of parameter uncertainties compared to the traditional method. Our method rests on cutoff frequencies, and hence decorrelation times of the overturning, and thereby introduces the unit of “time” into the analysis, whereas the “traditional” scheme relies on the total measured overturning which is not in itself time-dependent. Therefore, the focus of sensitivity in parameter space is different for both methods.

What can be obtained from consecutive measurements without a transformation back to the “real world” is a prediction of a trend in the bifurcation parameter. This may be worth something already, even if the absolute distance to the bifurcation point contains a certain margin of error.

A problem appears if the actual threshold in the climate system is not determined by the advective processes described by the Stommel model, but by a shutdown of convection before the saddle-node bifurcation is reached. If this is true, we can still estimate the distance from the saddle-node bifurcation, but this will not be relevant for the climate system. An estimate of the distance to the point where convection shuts down may also be possible, but this has not been investigated so far.

This paper illustrates that the variability of climatic variables contains important information that is not accessible if only averaged values are considered. Therefore, a special emphasis could be put on the analysis of variability and spectral properties in future climate research.

Acknowledgements We wish to thank Till Kuhlbrodt and Sven Titz for the helpful discussions. We also thank Adam Monahan and an anonymous reviewer for their helpful comments, which greatly improved the manuscript. T.K. acknowledges support by the Volkswagen Foundation.

References

- Bond G, Showers W, Cheseby M, Lotti R, Almasi P, de-Menocal P, Priore P, Cullen H, Hajdas I, Bonani G (1997) A pervasive millennial-scale cycle in North Atlantic holocene and glacial climates. *Science* 278(5341): 1257–1266
- Bryan K, Hansen FC (1995) A stochastic model of North Atlantic climate variability on decade-to-century time scales. In: Martinson DG, Bryan K, Ghil M, Hall MM, Karl TR, Sarachik ES, Sorooshian S, Talley LD (eds). *Natural climate variability on decade-to-century time scales*. National Academy Press, Washington DC
- Cessi P (1994) A simple box model of stochastically thermohaline flow. *J Phys Oceanogr* 24(8): 1911–1920
- Clark PU, Pisias NG, Stocker TF, Weaver AJ (2002) The role of the thermohaline circulation in abrupt climate change. *Nature* 415: 863–869
- Deser C, Holland M, Reverdin G, Timlin M (2002) Decadal variations in Labrador sea-ice cover and North Atlantic sea-surface temperatures. *J Geophys Res* 107(C5):10.1029/2000JC000683
- Gardiner CW (1994) *Handbook of stochastic methods*, 2nd ed. Springer, Berlin Heidelberg New York
- Haken H (1980) *Synergetics*, 2nd ed. Springer, Berlin Heidelberg New York
- Hasselmann K (1976) Stochastic climate models, part I. Theory. *Tellus* 28: 473–485
- Houghton RW, Visbeck MH (2002) Quasidecadal salinity fluctuations in the Labrador Sea. *J Phys Oceanogr* 32(2): 687–701
- Kloeden PE, Platen E (1999) *Numerical solution of stochastic differential equations*, 3rd ed. Springer, Berlin Heidelberg New York
- Knutti R, Stocker TF (2002) Limited predictability of the future thermohaline circulation close to an instability threshold. *J Climate* 15(2): 179–186
- Manabe S, Stouffer RJ (1988) Two stable equilibria of a coupled ocean-atmosphere model. *J Climate* 1: 841–866
- Monahan AH (2002) Stabilization of climate regimes by noise in a simple model of the thermohaline circulation. *J Phys Oceanogr* 32(7): 2072–2085
- Monahan AH, Timmermann A, Lohmann G (2002) Comments on “noise-induced transitions in a simplified model of the thermohaline circulation”. *J Phys Oceanogr* 32(3): 1112–1116
- Rahmstorf S (1995) Bifurcations of the atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* 378: 145–149
- Rahmstorf S (1996) On the freshwater forcing and transport of the Atlantic thermohaline circulation. *Climate Dynamics* 12: 799–811
- Rahmstorf S, Marotzke J, Willebrand J (1996) Stability of the thermohaline circulation. In: Krauss W (ed) *The warm water sphere of the North Atlantic ocean*, Borntraeger, Stuttgart, pp 129–158
- Rahmstorf S, Ganopolski A (1999a) Long-term global warming scenarios computed with an efficient coupled climate model. *Climatic Change* 43: 353–367
- Rahmstorf S, Ganopolski A (1999b) Simple theoretical model may explain apparent climate instability. *J Climate* 12: 1349–1352
- Risken H (1996) *The Fokker–Planck equation*, 3rd ed. Springer, Berlin Heidelberg New York
- Rooth C (1982) Hydrology and ocean circulation. *Prog Oceanogr* 11: 131–149
- Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. *Nature* 413: 591–596
- Schiller A, Mikolajevicz U, Voss R (1997) The stability of the North Atlantic thermohaline circulation in a coupled ocean-atmosphere general circulation model. *Climate Dynamics* 13(5): 325–347
- Scott JR, Marotzke J, Stone PH (1999) Interhemispheric thermohaline circulation in a coupled box model. *J Phys Oceanogr* 29: 351–365

- Smith JB, Schellnhuber H, Mirza MQ (2001) Lines of evidence for vulnerability to climate change: a synthesis. In: Canziani O, Carthy JM (eds) *Climate change 2001: impacts, adaptation and vulnerability – contribution of working group II to the Third Assessment Report IPCC*, Chap 19, Cambridge University Press, Cambridge, pp 914–967
- Sornette D (2000) *Critical phenomena in natural sciences*. Springer, Berlin Heidelberg New York
- Stocker TF, Schmittner A (1997) Influence of CO₂ emission rates on the stability of the thermohaline circulation. *Nature* 388: 862–865
- Stommel H (1961) Thermohaline convection with two stable regimes of flow. *Tellus* 13: 224–241
- Stommel HM, Young WR (1993) The average T–S relation of a stochastically forced box model. *J Phys Oceanogr* 23: 151–158
- Timmermann A, Lohmann G (2000) Noise-induced transitions in a simplified model of the thermohaline circulation. *J Phys Oceanogr* 30(8): 1891–1900
- Titz S, Kuhlbrodt T, Rahmstorf S, Feudel U (2002) On freshwater-dependent bifurcations in box models of the interhemispheric thermohaline circulation. *Tellus* 54(1): 89–98
- Tziperman E (2000) Proximity of the present-day thermohaline circulation to an instability threshold. *J Phys Oceanogr* 30: 90–104
- von Storch J-S, Müller P, Stouffer RJ, Voss R, Tett SFB (2000) Variability of deep-ocean mass transport: spectral shapes and spatial scales. *J Climate* 13(11): 1916–1935
- Walsh JE, Portis DH (1999) Variations of precipitation and evaporation over the North Atlantic ocean, 1958–1997. *J Geophys Res* 104(D14): 16613–16631
- Warren BA (1981) Deep circulation of the world ocean. In: Warren BA, Wunsch C (eds). *Evolution of physical oceanography*. MIT Press, Cambridge MA, pp 6–40