Stability and Variability of Open-Ocean Deep Convection in Deterministic and Stochastic Simple Models

Dissertation

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
in der Wissenschaftsdisziplin Physik der Ozeane

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von

Till Kuhlbrodt

geboren am 27.1.1969 in Hamburg

Potsdam, im Juli 2002

Diseur de bons mots, mauvais caractère.

Blaise Pascal (1623–1662)

Are you confident about the physics of the
winter-time convection and the subduction
processes?

Henry Stommel, A View of the Sea (1987)

Abstract

Deep convection is an essential part of the circulation in the North Atlantic Ocean. It influences the northward heat transport achieved by the thermohaline circulation. Understanding its stability and variability is therefore necessary for assessing climatic changes in the area of the North Atlantic.

This thesis aims at improving the conceptual understanding of the stability and variability of deep convection. Observational data from the Labrador Sea show phases with and without deep convection. A simple two-box model is fitted to these data. The results suggest that the Labrador Sea has two coexisting stable states, one with regular deep convection and one without deep convection. This bistability arises from a positive salinity feedback that is due to the net freshwater input into the surface layer. The convecting state can easily become unstable if the mean forcing shifts to warmer or less saline conditions.

The weather-induced variability of the external forcing is included into the box model by adding a stochastic forcing term. It turns out that deep convection is then switched "on" and "off" frequently. The mean residence time in either state is a measure of its stochastic stability. The stochastic stability depends smoothly on the forcing parameters, in contrast to the deterministic (non-stochastic) stability which may change abruptly. The mean and the variance of the stochastic forcing both have an impact on the frequency of deep convection. For instance, a decline in convection frequency due to a surface freshening may be compensated for by an increased heat flux variability.

With a further simplified box model some stochastic stability features are studied analytically. A new effect is described, called wandering monostability: even if deep convection is not a stable state due to changed forcing parameters, the stochastic forcing can still trigger convection events frequently. The analytical expressions explicitly show how wandering monostability and other effects depend on the model parameters. This dependence is always exponential for the mean residence times, but for the probability of long nonconvecting phases it is exponential only if this probability is small. It is to be expected that wandering monostability is relevant in other parts of the climate system as well.

All in all, the results demonstrate that the stability of deep convection in the Labrador Sea reacts very sensitively to the forcing. The presence of variability is crucial for understanding this sensitivity. Small changes in the forcing can already significantly lower the frequency of deep convection events, which presumably strongly affects the regional climate.

Kurzzusammenfassung

Die Tiefenkonvektion ist ein wesentlicher Bestandteil der Zirkulation im Nordatlantik. Sie beeinflusst den nordwärtigen Wärmetransport der thermohalinen Zirkulation. Ein Verständnis ihrer Stabilität und Variabilität ist daher nötig, um Klimaveränderungen im Bereich des Nordatlantiks einschätzen zu können.

Diese Arbeit hat zum Ziel, das konzeptionelle Verständnis der Stabilität und der Variabilität der Tiefenkonvektion zu verbessern. Beobachtungsdaten aus der Labradorsee zeigen Phasen mit und ohne Tiefenkonvektion. Ein einfaches Modell mit zwei Boxen wird an diese Daten angepasst. Das Ergebnis legt nahe, dass die Labradorsee zwei koexistierende stabile Zustände hat, einen mit regelmäßiger Tiefenkonvektion und einen ohne Tiefenkonvektion. Diese Bistabilität ergibt sich aus einer positiven Salzgehalts-Rückkopplung, deren Ursache ein Netto-Süßwassereintrag in die Deckschicht ist. Der konvektive Zustand kann schnell instabil werden, wenn der mittlere Antrieb sich hin zu wärmeren oder weniger salzhaltigen Bedingungen ändert.

Die wetterbedingte Variabilität des externen Antriebs wird durch die Addition eines stochastischen Antriebsterms in das Modell eingebaut. Es zeigt sich, dass dann die Tiefenkonvektion häufig an- und wieder ausgeschaltet wird. Die mittlere Aufenthaltszeit in beiden Zuständen ist ein Maß ihrer stochastischen Stabilität. Die stochastische Stabilität hängt in glatter Weise von den Parametern des Antriebs ab, im Gegensatz zu der deterministischen (nichtstochastischen) Stabilität, die sich abrupt ändern kann. Sowohl das Mittel als auch die Varianz des stochastischen Antriebs beeinflussen die Häufigkeit von Tiefenkonvektion. Eine Abnahme der Konvektionshäufigkeit, als Reaktion auf eine Abnahme des Salzgehalts an der Oberfläche, kann zum Beispiel durch eine Zunahme der Variabilität in den Wärmeflüssen kompensiert werden.

Mit einem weiter vereinfachten Box-Modell werden einige Eigenschaften der stochastischen Stabilität analytisch untersucht. Es wird ein neuer Effekt beschrieben, die wandernde Monostabilität: Auch wenn die Tiefenkonvektion aufgrund geänderter Parameter des Antriebs kein stabiler Zustand mehr ist, kann der stochastische Antrieb immer noch häufig Konvektionsereignisse auslösen. Die analytischen Gleichungen zeigen explizit, wie die wandernde Monostabilität sowie andere Effekte von den Modellparametern abhängen. Diese Abhängigkeit ist für die mittleren Aufenthaltszeiten immer exponentiell, für die Wahrscheinlichkeit langer nichtkonvektiver Phasen dagegen nur dann, wenn diese Wahrscheinlichkeit gering ist. Es ist zu erwarten, dass wandernde Monostabilität auch in anderen Teilen des Klimasystems eine Rolle spielt.

Insgesamt zeigen die Ergebnisse, dass die Stabilität der Tiefenkonvektion in der Labradorsee sehr empfindlich auf den Antrieb reagiert. Die Rolle der Variabilität ist entscheidend für ein Verständnis dieser Empfindlichkeit. Kleine Änderungen im Antrieb können bereits die Häufigkeit von Tiefenkonvektionsereignissen deutlich mindern, was sich vermutlich stark auf das regionale Klima auswirkt.

Contents

Chapter 1

Introduction

1.1 Deep convection as a part of the global ocean circulation

Most of the processes that make up the ocean circulation are continuous over long periods of time and extend over large parts of the oceans. The surface fluxes of heat and freshwater cause the deep overturning circulation known as the thermohaline circulation (THC). Eddies that form in regions of strong horizontal density gradients mix the waters laterally. The momentum transferred by the winds affects the upper layer of the ocean: it induces a shallow vertical mixing, and it drives a large-scale advective circulation.

Open-ocean deep convection differs from all those processes. Deep convection events last but a few days and occur only in a few small areas. These areas are marked by a particularly weak vertical density gradient, in contrast to the strong vertical density stratification that is found throughout most of the world ocean. Deep convection happens when strong surface forcing, like rapid cooling in winter, is capable of increasing the density in the upper layer to a point where the vertical gradient vanishes. In a few days, the convecting water column is then vigorously mixed throughout its volume of, typically, some ten kilometres in diameter and about two kilometres depth. A comprehensive review of the physics and phenomenology of deep convection was given by Marshall and Schott (1999).

Why the interest in deep convection? In spite of its intermittent and scarce appearance, it is a crucial part of the global circulation for two reasons: it triggers the deep-water formation that is an essential part of the THC, and it is a bottle-neck for the global heat fluxes in the ocean.

Concerning the deep-water formation first, one finds that deep convection occurs in those areas that have the highest surface density of the world ocean. In the North Atlantic, these are the Greenland-Iceland-Norwegian Sea and the central Labrador Sea (Fig. 1.1). After a deep convection event, the very dense, convectively mixed water slowly sinks to depth and entrains

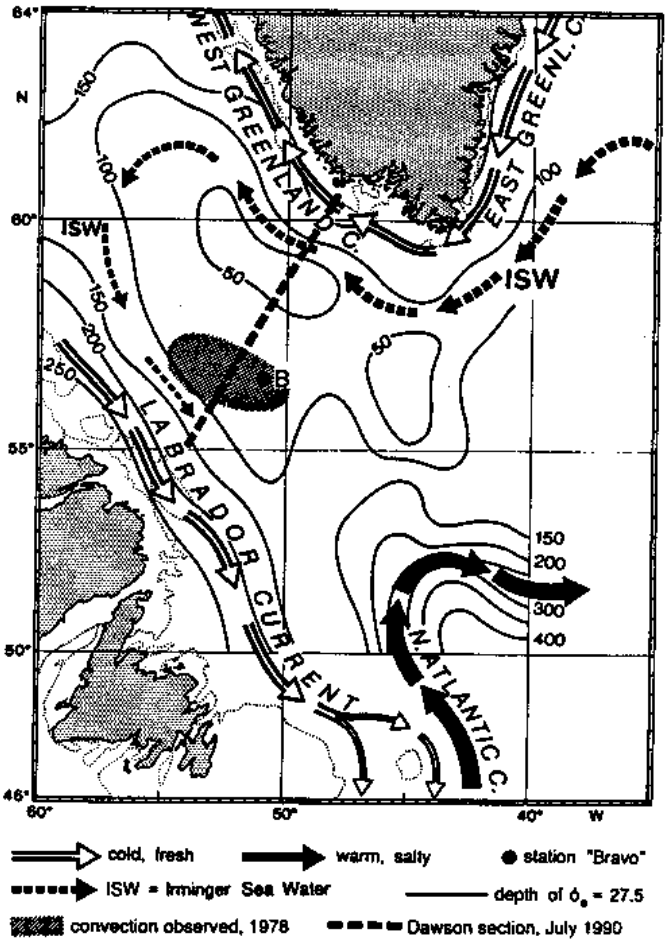

Figure 1.1: *Map of the central Labrador Sea between the southern tip of Greenland (top margin) and Newfoundland (lower left corner). Arrows show the mean circulation (neglecting eddies): white for surface currents, dashed and black for intermediate depth currents. In contrast to the strong boundary currents there is only a very weak mean velocity in the interior Labrador Sea. The position of Ocean Weather Ship Bravo referred to in chapter 2 is denoted by "B" (in the hatched deep convection area). The map is from Marshall and Schott (1999).*

the neighbouring waters. The horizontal density gradient between these dense waters in the Northern Atlantic and the lighter waters at lower latitudes and in the Southern Atlantic drives the THC. The dense waters flow southwards at depth, and at the same time there is a northward flow of warmer and lighter surface waters. This flow transports huge amounts of heat northwards (Roemmich and Wunsch, 1985; Ganachaud and Wunsch,

2000), an indispensable contribution to the relatively mild climate in Europe (Rahmstorf and Ganopolski, 1999).

Concerning the role of deep convection for the ocean's heat budget, one observes that during deep convection events the deep ocean loses enormous amounts of heat to the atmosphere in short periods of time. This rapid small-scale heat loss balances the slow heat gain of the deep ocean due to various large-scale mixing processes (Kunze and Sandford, 1996; Munk and Wunsch, 1998).

Apart from being a spectacular physical process of its own, the twofold relevance of deep convection for the global ocean circulation calls for a deepened understanding of its stability and variability. The present work contributes to this understanding by addressing *three guiding questions:*

Guiding Questions

GQ1: How sensitive is deep convection in the Labrador Sea to changes in climate?

GQ2: Defining "stability" as "sensitivity to perturbations" — what role does external climate variability play in determining the stability of deep convection?

GQ3: How can this role be explained theoretically in a simple conceptual way?

The following sections of this Introduction sketch our strategy to tackle these questions, and they review the relevant and recent research.

1.2 Observations of open-ocean deep convection and its variability

Direct observations of deep convection events are rare due to their small extent in space and time. However, the cold and dense water mass that deep convection leaves behind (Lazier, 1973; Send and Marshall, 1995) is easily detectable many months later. The data from the past decades (Lazier, 1980, 1988; Dickson et al., 1988, 1996; Belkin et al., 1998; Khatiwala et al., 2002) display a great interannual to decadal variability in the occurrence and depth of deep convection events in the Labrador Sea. Phases with deep convection occurring regularly each winter alternate with phases where deep convection is absent. Our focus is on the Labrador Sea because much less data are available from other deep convection sites.

The observations suggest that deep convection is very sensitive to changes in the forcing of the upper layer of the water column. The main

contributors to the buoyancy forcing of the interior Labrador Sea are the local surface flux of heat and the lateral freshwater transport (Marshall and Schott, 1999). The local surface heat fluxes are determined by the weather conditions (Lab Sea Group, 1998; Lilly et al., 1999). Thus, the interannual variability of the strength of deep convection is strongly correlated with the leading interannual variability mode of the atmosphere over the North Atlantic, the North Atlantic Oscillation (Dickson et al., 1996). The freshwater forcing of the interior Labrador Sea operates in two steps: freshwater from sea-ice melt and runoff is advected by the boundary currents (Fig. 1.1), and then lateral eddy mixing transports the freshwater into the central Labrador Sea (Dickson et al., 1988, 1996; Houghton and Visbeck, 2002). The sensitivity of deep convection is also inferred from its apparent absence from the Labrador Sea under different climatic conditions in the past (Hillaire-Marcel et al., 2001).

A prominent feature of the variability in the North Atlantic are Great Salinity Anomalies (GSAs). During a GSA a pool of anomalously fresh water travels through the North Atlantic (Dickson et al., 1988; Belkin et al., 1998). The arrival of such a freshwater anomaly weakens or even stops deep convection in the Labrador Sea. Particularly well-documented is the GSA that suppressed deep convection in the years 1969–1971 (Lazier, 1980).

It is the first step of our research strategy to develop a conceptual approach that is directly based on Lazier's observations. The mean vertical profile in the Labrador Sea suggests a two-box approach. The parameters of the two-box model are estimated from a fit to Lazier's data. The model is then successful in reproducing the observed evolution (chapter 2), particularly the switches between the convecting and the nonconvecting phases.

Analyses of the hydrographic and meteorological observations (Dickson et al., 1996; Lilly et al., 1999) revealed that several processes act together to stop convection: apart from the lateral freshwater flux, these are weak local heat flux forcing at the surface, weak cyclonic wind stress (too weak to decrease the density stratification of the water column), and a local amplifying feedback, the *positive salinity feedback* (Fig. 1.2). This feedback arises from the annual mean freshwater input into the upper layer. As long as deep convection occurs regularly every winter, the freshwater input is removed by the mixing with the more saline deep water mass. If deep convection is however suppressed, maybe in an anomalously warm winter, the freshwater accumulates. The surface layer becomes less and less dense, strengthening the vertical density gradient. The longer the absence of deep convection lasts, the stronger a forcing must be that can overcome the vertical density gradient, restart deep convection, and mix the water column again.

It is not fully settled yet how the heat and freshwater forcing interact with the positive salinity feedback in triggering and maintaining a phase of absent deep convection. Previous studies (Dickson et al., 1996; Lilly et al.,

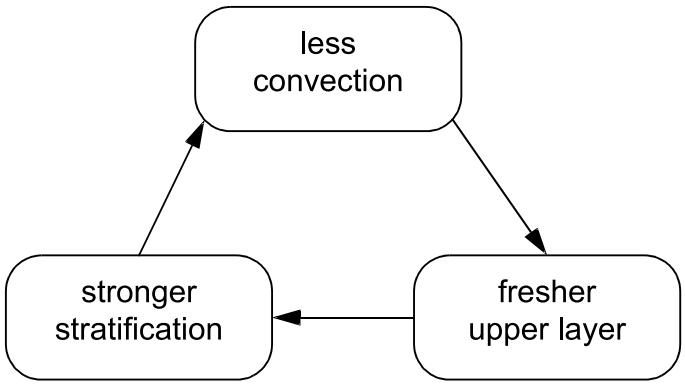

Figure 1.2: *The positive salinity feedback.*

1999) only assume a passive ocean that merely reacts to the external forcing, where the local positive salinity feedback plays a minor role.

The positive salinity feedback is easily captured by our two-box model through an appropriate choice of the boundary conditions. Thus, our conceptual model is an ideal tool to formulate our main hypothesis:

Main Hypothesis

Deep convection is interrupted by a short anomaly in the surface forcing, either in the heat fluxes or in the lateral freshwater fluxes. After such an anomaly the local positive salinity feedback actively works to enhance the vertical density gradient further.

The interplay of the forcing anomalies and the positive salinity feedback proposes a mechanism for the observed switches between convecting and nonconvecting phases. Moreover it explains why deep convection in the Labrador Sea is so sensitive to the variability of different forcings. With the model parameters estimated from the observational data we can quantify this sensitivity (ch. 2; also Kuhlbrodt et al. (2001)), which is one part of the answer to the first guiding question (GQ1).

To assume a positive salinity feedback being active in the phases of absent deep convection contrasts with the conclusions of Dickson et al. (1996) and Lilly et al. (1999). Our results however will show that this assumption is well justified. We stress that in a very recent observational study (Houghton and Visbeck, 2002) the positive salinity feedback was employed to close the Labrador Sea freshwater budget.

1.3 The stability of deep convection in models

The concept of the *stability* of a climate or model state is crucial for this study. We call a state *stable* under two conditions: (i) it is an equilibrium state, meaning that it does not change in time (apart from the seasonal cycle), and (ii), if small perturbations are applied, they are not amplified, but eventually the equilibrium state is reached again. This stability definition is sufficient for the first parts of the study (ch. 2). Later on, we will add a quantitative aspect, asking how strong the perturbations must be to leave the stable state (ch. 3). A mathematically more rigorous definition will be provided in chapter 4. Note that the dynamic stability we deal with here is not to be confused with the static stability of a stratified fluid.

How are deep convection and its stability represented in models? High-resolution, three-dimensional ocean general circulation models (OGCMs) capture the large-scale features of deep convection. For instance, the intensity and location of deep convection influence the northward heat transport of the THC (Rahmstorf, 1995a; Häkkinen, 1999) and the total meridional overturning (Rahmstorf, 1995c; Böning et al., 1996). However, due to their small spatial extent and their short duration the convection events themselves are not resolved in most of the OGCMs. For this reason they have to be parameterized.

Vertical mixing depends in a highly nonlinear way on the vertical density gradient. It is very weak for a strong vertical density gradient, but if this density gradient approaches zero, the vertical mixing rapidly intensifies, becoming the vigorous convective mixing. In most models this nonlinearity is parameterized in its step-function limit, which is dubbed convective adjustment (CA). Wherever there is a grid box with denser water overlying a grid box with lighter water, the CA schemes (Marotzke, 1991; Rahmstorf, 1993; Klinger et al., 1996) remove this static instability in the water column by completely mixing the waters in such vertically adjacent boxes. These schemes have some drawbacks: for instance, the convective mixing may be too intense (Lilly et al., 1999), or grid-scale instabilities may occur (Cessi, 1996b; Molemaker and Dijkstra, 2000). Nevertheless, given the need to minimize computing time for OGCMs, CA proved a sufficient parameterization for many applications.

A simple conceptual model of deep convection consists of two boxes, one for the permanently mixed surface layer and one for the deep ocean (Welander, 1982). It includes different boundary conditions for the surface fluxes of heat and freshwater. Convective mixing is modeled by CA. Under boundary conditions that mimic the positive salinity feedback, the nonlinearity of the convective mixing may lead to a bistability of the water column where convection can be either permanently "on" or permanently "off", depending only on the initial condition. The point of zero density stratification is then a threshold that separates two model regimes, a convecting one and a

nonconvecting one. They correspond to the phases of regular or absent deep convection seen in the observational data.

Using Welander's box model, Lenderink and Haarsma (1994, 1996) showed that large regions of a simple model ocean are indeed bistable in this sense. With small salinity perturbations at grid points in those regions, convection can be switched "off" or "on". Often this leads to convection switches in neighbouring grid points, and eventually the whole circulation pattern of the thermohaline overturning reorganizes. In this respect the THC exhibits multistability: small anomalies applied under constant boundary conditions make the model jump between different circulation states. Rahmstorf (1994, 1995c), Pierce et al. (1995), and Hirschi et al. (1999) studied the same kind of multistability in idealized OGCMs. It appears as well in an OGCM with realistic topography (Rahmstorf, 1995a,b). Convective multistability is also a function of the spatial resolution of an OGCM and its convection parameterization (Cessi and Young, 1996; Vellinga, 1998; Molemaker and Dijkstra, 2000).

From all those studies it is obvious that the switching mechanism of our main hypothesis is relevant in OGCMs and in box models. We want to address the open question whether it operates in the real ocean, too (ch. 2). The box model we use is based on Rahmstorf's (2001) extended version of Welander's (1982) model. We will show that the Labrador Sea is indeed bistable with respect to deep convection, also in the presence of the seasonal cycle.

Following our main hypothesis, we apply single anomalies in the forcing to make the model switch between the two states. Our results demonstrate that with this mechanism the model quite satisfyingly reproduces the two convection switches seen in Lazier's (1980) data. This renders our main hypothesis fairly plausible, and lays a basis for the further work.

The next issue addressed in chapter 2 is the sensitivity of both the convecting and the nonconvecting state to changes in the mean heat and freshwater fluxes. We will show that small changes in these fluxes may destabilize the convecting state. This high sensitivity is one part of the answer to the first guiding question (GQ1).

1.4 Stability in the presence of variability

The variability of the surface weather forcing and the freshwater fluxes strongly influence the convective activity in the Labrador Sea (section 1.2). How to include this external variability in the box model in a possibly simple way? In order to address GQ2 we need to extend our model. So far it has been *deterministic:* apart from single, arbitrary anomalies the forcing terms do not include any variability.

What difference does the inclusion of climate variability make in GCMs? In general, OGCMs have a climatological atmospheric boundary condition. They are designed to model a mean circulation state, but not the variability of the circulation. Having studied the high sensitivity of the deep water formation areas in their OGCMs, Rahmstorf (1995c) and Lenderink and Haarsma (1996) already pointed out the need to study the convective multistability under the influence of the seasonal cycle and of stochastic forcing.

In contrast to OGCMs, coupled GCMs (CGCMs) include a fully dynamical, variable atmosphere. Some CGCMs actually show how deep convection in the Labrador Sea switches "on" and "off" (Tett et al. (1997); P. Wu, pers. comm.), and how such convective variability is linked with variability in the overall THC overturning (Delworth et al., 1993; Cooper and Gordon, 2002). The convective variability is obviously part of the equilibrium climate state in those models.

In some global warming runs conducted with CGCMs one sees that deep convection in the Labrador Sea strongly weakens in the early 21st century (Flato and Boer, 2001) or even completely ceases (Wood et al., 1999). Such a final shutdown of deep convection goes along with strong regional climate changes: the atmosphere lacks the convective heat release from the ocean and cools. Hence, the understanding of the stability of deep convection in the present climate is essential for an assessment of possible future climatic changes in the North Atlantic area.

A simple conceptual way to introduce climate variability in the box model is to add Gaussian noise to the forcing terms in the model equations. This so-called *stochastic climate model* was proposed by Hasselmann (1976) and applied to sea-surface temperatures by Frankignoul and Hasselmann (1977). Hall and Manabe (1997) showed that it works as well for sea surface salinities. Hasselmann's approach is explained in detail in chapter 3. Since its early days, stochastic forcing has been used in many studies to drive OGCMs (Weisse et al., 1994; Skagseth and Mork, 1998), climate models of intermediate complexity (Mysak et al., 1993; Aeberhardt et al., 2000; Knutti and Stocker, 2002; Ganopolski and Rahmstorf, 2002), and ocean box models (Stommel and Young, 1993; Cessi, 1994, 1996a; Griffies and Tziperman, 1995; Pierce et al., 1995; Lohmann and Schneider, 1999; Timmermann and Lohmann, 2000; Monahan, 2002).

Here we use a stochastic term to mimic weather variability (ch. 3), and we model freshwater anomalies stochastically, too (ch. 4), taking up an early suggestion of Dickson et al. (1988). The basic concepts of Hasselmann's (1976) stochastic climate model and Welander's (1982) two-box model are thus combined to a *stochastic* deep convection box model.

In contrast to the deterministic model used in chapter 2, anomalies are now always present. In the bistable domain this leads to frequent jumps between the two model states of convection occurring regularly ("on") or

not at all ("off"). This provides a simple mechanism of how the convection "flickering" seen in CGCMs might come about.

It will turn out that we need a different definition of stability if stochastic variability is present. The theory of random dynamical systems (Freidlin and Wentzell, 1998) provides that definition. The average time that the model trajectory stays in one state before it jumps away again, the *mean residence time*, is a convenient measure of stochastic stability. In this way, the variability is exploited to measure the stability. We use the mean residence time to quantitatively compare the stability of the convecting and the nonconvecting state. This concept of quantitative stability does not exist in the deterministic picture. Thus there are two approaches to answer GQ1: the *deterministic* approach (ch. 2), and the *stochastic* approach (ch. 3).

We will see that the frequent jumps between the model states occur as well if one of the two model states is not stable. Changing a forcing parameter may lead to an abrupt stability loss in the deterministic model, but in the stochastic model this induces only a smooth decrease in the frequency of jumps. This is the main answer to GQ2 (ch. 3; see also Kuhlbrodt et al. (2001)): in the stochastic model that includes climate variability, switches between convecting and nonconvecting phases happen regularly. The parameters of the deterministic model determine only how often these jumps occur. In fact, similar results have been obtained with a box model of the THC (Cessi, 1994) and with a box model of the Welander type mimicking oscillatory states of an OGCM (Cessi, 1996a), but yet no study has addressed the bistability of deep convection derived from observations.

Another issue dealt with in chapter 3 are the distributions of the residence times. They reveal the influence of the positive salinity feedback in the stochastic model.

In order to obtain reliable statistics of the residence times many long model runs are needed. Working with a simple model is a major advantage here because a GCM is computationally far too expensive for this purpose. Another major advantage of simple models is that only the most relevant processes are included. While a simple model is open to thorough physical understanding, it is often difficult to isolate the main processes and feedbacks in a GCM. Since our focus is on the conceptual and theoretical understanding, studies with more comprehensive models (like two-dimensional models of the Labrador Sea (Visbeck et al., 1997; Khatiwala and Visbeck, 2000), or OGCMs) have been left for future research.

1.5 Understanding stochastic stability

Seeing the pronounced differences between deterministic and stochastic stability, it is our aim to explain them in the simplest possible way. For this purpose we want to apply the existing theory on simple stochastic dynami-

cal systems (see the textbooks of Freidlin and Wentzell (1998) and Gardiner (2002)). In particular, there is a large body of work on mean residence times that has been employed in many branches of physics (Hänggi et al., 1990). Freidlin and Wentzell's concept of stochastic stability, outlined in chapter 4, relies on mean residence times. If we want to understand stochastic stability *analytically* in our model, we need a strongly simplified version.

In contrast to the familiar non-stochastic (or deterministic) stability theory, the theory of stochastic stability has not often been used in ocean dynamics. The papers of Cessi (1994, 1996a) have already been mentioned. Timmermann and Lohmann (2000) suggested that multiplicative noise (whose strength is a function of the variables) may excite additional stochastically stable model states, but this suggestion was erroneous (Monahan et al., 2002). Very recently, Monahan (2002) worked out a number of differences between deterministic and stochastic stability using a bistable THC box model.

Climate change may manifest itself as a changing preference of specific climate regimes (Houghton et al., 2001). A recently observed trend to strong deep convection in the Labrador Sea (Dickson et al., 2002) might be interpreted in this way. Following the idea of Palmer (1999), Monahan (2002) as well as Khatiwala et al. (2001) showed that from simple stochastic bistable models one can learn how this regime preference works, and on which parameters it depends. Yet, there is still need for a deeper conceptual understanding of the stability of climate regimes. In addressing the third of our guiding questions (GQ3) we strive for contributing to this fundamental issue.

It will turn out that a simple stochastic model of two potential wells separated by the convection threshold is sufficient to capture the stability properties of the two-box model (ch. 4). This strongly simplified model still has the main physical ingredients: the convection threshold mimics the nonlinearity of the vertical mixing, and by giving the two wells differing sizes the positive salinity feedback is represented.

With this simplest possible model of deep convection we can see analytically how the residence times depend on the climate parameters. It is thus the proper tool to address GQ3. In addition, by analyzing the distribution of the residence times we can state the results of Khatiwala et al. (2001) more precisely.

We will show in chapter 4 that the stochastic forcing has two contrasting effects on the stability. On the one hand, the stochastic model often occupies only one of two potential wells, as described by Monahan (2002). On the other hand, if only the well of the nonconvecting state is present, convection events may still be triggered frequently. We will call this new effect wandering monostability, because the model trajectory wanders through both model regimes even if only the nonconvecting state is stable. Eventually this

might be the core process behind the deep convection variability observed in the Labrador Sea.

In the subsequent chapter 2 the observational data are analyzed, and with the two-box model the stability of deep convection is studied (GQ1). The crucial effect of variability on the stability is explored in chapter 3 (GQ1, GQ2), along with some quantitative aspects of the variability. With the theoretical underpinning developed in chapter 4 that effect is explained in an analytical way (GQ3). Finally, the summarizing chapter 5 gives the overall view on the results and points out some lines of future research.

Chapter 2

Deterministic stability analysis of Labrador Sea deep convection

2.1 Overview

In this chapter a deterministic (non-stochastic) approach to model deep convection in a conceptual way is followed. A detailed analysis of observational data from one of the sites where deep convection occurs in the North Atlantic, namely the Labrador Sea (section 2.2), highlights the differences between the behaviour of the surface mixed layer and the deep interior of the ocean. This is the motivation for developing a conceptual two-box model (section 2.3). Since the variables are temperature T and salinity S in each of the 2 boxes, it is dubbed the 2TS model. Previous work with similar models is extended by fully including the seasonal cycle in the surface forcing. This extension opens up the possibility to estimate the model parameters from the observational data (section 2.4). The observations show marked differences between periods with and without deep convection. This is reflected in the box model by two possible stable states: one convecting and one non-convecting. The two states overlap in the parameter space; this bistability is studied in section 2.5. In addition, we are able to locate Labrador Sea deep convection – as represented by the observations – in a stability diagram, which allows some inferences regarding its stability in a changing climate. These are found in the concluding section 2.6.

2.2 Observational data from OWS Bravo

Long time series of hydrographic data that show clear signs of deep convection events are rare. The data from Ocean Weather Ship (OWS) Bravo are exceptional due to their location and their sampling rate. OWS Bravo was

located in the central Labrador Sea close to the area of the deepest convection events (Fig. 1.1). From January 1964 through September 1974 the sampling rate of the data varied between six hours and two months. This enables the derivation of a time series of monthly means that clearly reflects the winter open ocean deep convection events. Other time series that extend longer in time (e.g. by the Canadian Department of Fisheries and Oceans, http://www.mar.dfo-mpo.gc.ca/science/ocean/woce/welcome.html) do not resolve the seasonal cycle in most of the years. Hence, a model fit to the data as carried out in section 2.4 would not be possible.

The original data (Lazier, 1980) were interpolated to standard depth levels. Potential temperature (θ) and potential density (σ_0) were computed with the standard formulae (Fofonoff and Millard Jr., 1984). To obtain monthly mean values, the data of each month were binned and averaged at each depth level. Missing monthly means were interpolated linearly. Subsequently, the data were averaged over the upper layer (0–50 m) and the deep layer (200–2000 m). The intermediate level (50–200 m) was left out because on the one hand this layer still shows substantial seasonal variations, but on the other hand it is not part of the surface mixed layer throughout the year. The next chapters will show that our results do not depend on the exact layer depths.

The resulting time series of monthly means are given in Fig. 2.1. The winters 1969–1971 show the impact of the Great Salinity Anomaly (GSA, described by Dickson et al. (1988)) that suppressed deep convection in the Labrador Sea by the advection of a large freshwater anomaly. In consequence, temperatures and salinities follow different trends in both layers: cooling and freshening in the upper layer lead to less dense waters, while the deep ocean is becoming slightly warmer and saltier. The upper layer values show a strong seasonal cycle. A small potential density difference between the two layers indicates deep convection. For a number of reasons, this difference is not exactly zero: the mixing of the layers occurs only during a few days, but the data are monthly averaged; the mixing does not occur necessarily exactly at the ship site and throughout the whole water column. In some winters the upper layer overshoots in temperature: to compensate for the deep ocean being saltier than the surface waters, the upper layer must become colder than the deep ocean before the vertical density gradient vanishes and deep convection starts.

In order to consider seasonal and interannual variability separately, we first computed the mean seasonal cycle of the time series (Fig. 2.2). In the upper layer, the temperature cycle has its minimum in February and its maximum in September, with an amplitude of 2.2°C. The salinity cycle lags by about one month: the cycle, with an amplitude of 0.13 psu, peaks in March and reaches its minimum in October. The σ_0 cycle lies in between with an amplitude of 0.29 kg m^{-3}. The deep layer seasonal cycles (not shown) are almost two orders of magnitude smaller. The deep layer is coldest

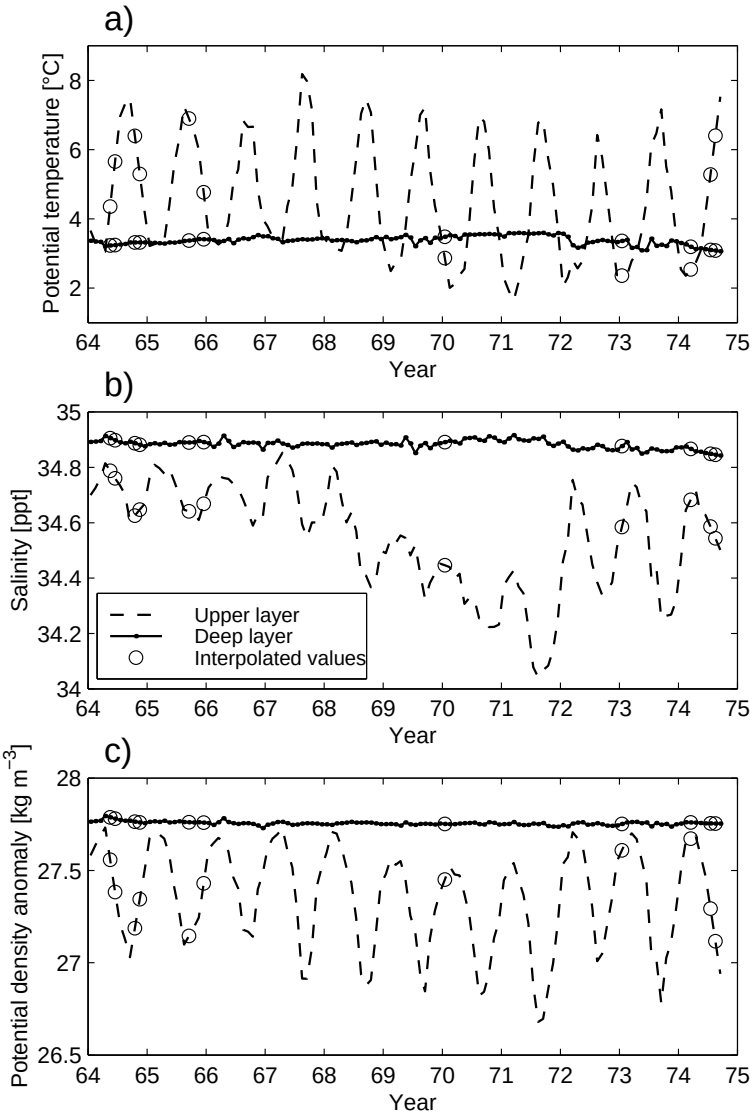

Figure 2.1: *Time series of monthly means obtained from the OWS Bravo data set: potential temperature (a), salinity (b), and potential density (c) of the upper layer (dashed) and the deep layer (dots). Interpolated values are indicated by circles. The large minimum density difference in the winters from 1969 to 1971 is an indication for the absence of deep convection, which led to the cooling and freshening of the upper layer.*

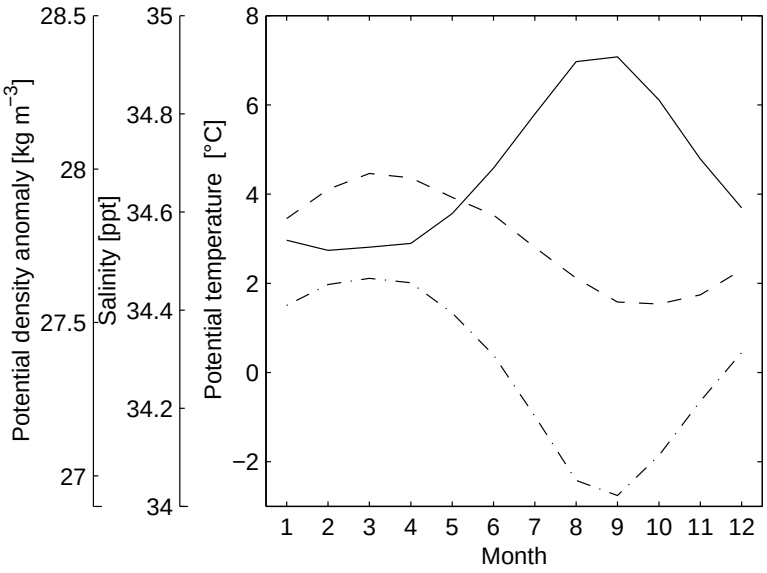

Figure 2.2: *Mean seasonal cycle of the upper layer (0–50 m) from OWS Bravo data for potential temperature (solid), salinity (dashed) and potential density (dash-dotted).*

and densest in April after the convection season. Afterwards there is a warming of 0.12°C and a density decrease of 0.015 kg m^{-3}, both of which reach their extremum in November/December. Deep layer salinity variations are very small.

The temperature cycle in the upper layer is mostly forced by fluxes of latent and sensible heat in winter and short-wave radiation in summer (Smith and Dobson, 1984). Various freshwater sources of inexactly known strength (Canadian runoff, melt-water, local precipitation, and low-salinity inflow from the Arctic Ocean) account for the salinity cycle (Lilly et al., 1999). Results from recent analyses that make use of several tracers (Khatiwala et al., 1999, 2002) try to quantify those different freshwater sources. The role of precipitation is apparently not large. Advected sea-ice that melts locally does not contribute strongly either. The main source of freshwater in the central Labrador Sea is through lateral eddy mixing from the boundary currents. While the seasonal cycle is mainly caused by sea-ice melt freshwater that is advected with the boundary currents, the net freshwater transport can be attributed to Arctic runoff being advected in the same way.

Next, we subtracted the seasonal cycle from each time series in Fig. 2.1. The resulting time series (Fig. 2.3) show the variability excluding the seasonal cycle. Three distinct phases stand out, marked by either the occurrence or the absence of convection. In the first convective phase (Phase 1), from January 1964 to March 1968, the values in the upper box fluctuate

with hardly any interannual trend. However, there are trends in the deep layer. Time series of single depth levels reveal that these trends are more pronounced in deeper layers. With maximum convection depth varying from year to year, the deeper layers sometimes remain untouched and accumulate heat over more than one year.

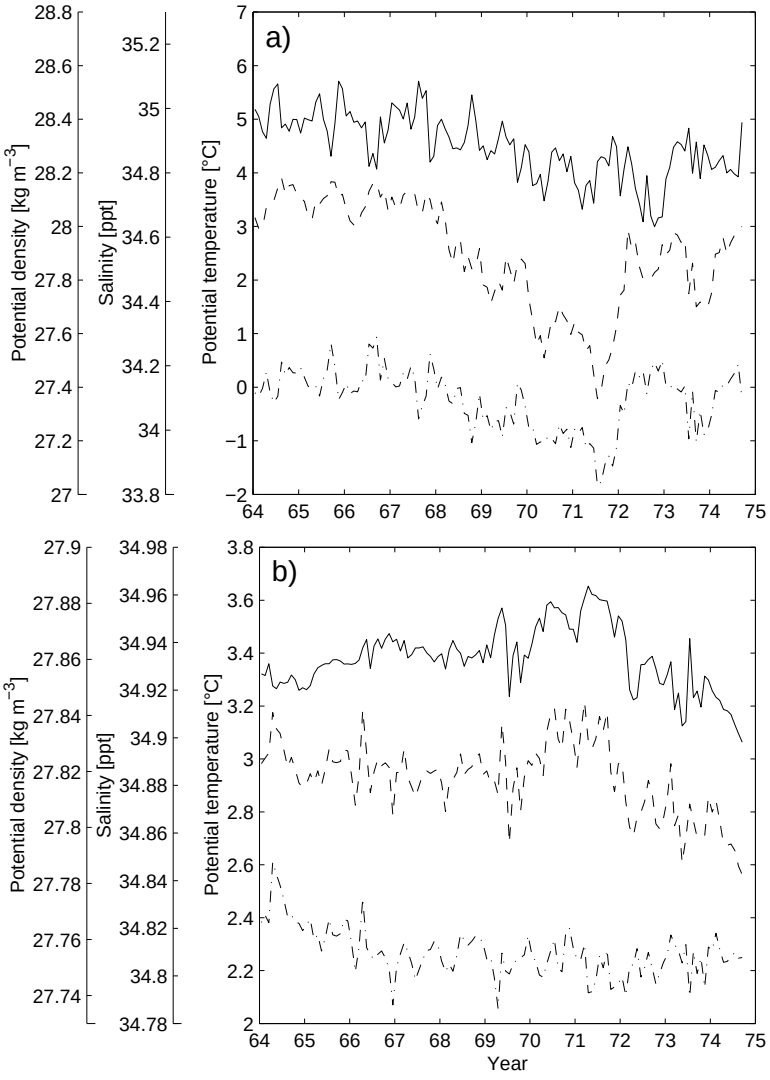

Figure 2.3: *Time series of monthly means with subtracted seasonal cycle for the upper layer (a) and the deep layer (b) of potential temperature (solid), salinity (dashed) and potential density (dash-dotted).*

The phase from April 1968 to September 1971 (Phase 2) is characterized by the passage of the GSA through the Labrador Sea, suppressing convection. Phase 2 begins after the last convection event and ends with the upper

Quantity	θ_1	S_1	$\sigma_{0,1}$
Trend (yr^{-1})	$-0.28°\mathrm{C}$	-0.11 psu	-0.060 kg m^{-3}
Quantity	θ_2	S_2	$\sigma_{0,2}$
Trend (yr^{-1})	$-0.070°\mathrm{C}$	-0.0072 psu	-0.0013 kg m^{-3}

Table 2.1: *Trends of potential temperature, salinity and potential density in the upper layer (index 1) and the deep layer (index 2) during the GSA (04/68 to 09/71) from the time series without seasonal cycle depicted in Figure 2.3.*

layer salinity starting to rise again. Annual trends of all quantities during Phase 2 are given in Table 2.1. Starting in 1968, a large amount of fresh surface water, advected by the Greenland current and then laterally mixed into the central Labrador Sea (Dickson et al., 1988), lowers the salinity in the upper layer. This suppresses deep convection; the absence of convection leads to further cooling and freshening and a density decrease in the upper layer. While the cooling comes to a halt already in early 1970, the strong freshening continues until late 1971. Our view on the role of the GSA freshwater anomaly in the Labrador Sea is different from the view of Dickson et al. (1988) and Lilly et al. (1999). While they assume that the freshwater anomaly lasts four years and thus inhibits convection, it is our hypothesis that a short anomaly suppresses convection in one winter, and that the evolution in the following years is due to the local positive salinity feedback.

In the deep layer, the waters become slightly warmer and more saline during Phase 2, mostly by lateral mixing from adjacent water masses. Potential density in the deep layer shows a hardly significant decreasing trend that is clearly smaller than in Phase 1, as the warming and salinification partly compensate in their effect on density.

Phase 3 is again characterized by annual convection events. Starting in October 1971, strong wind mixing of the surface mixed layer caused it to deepen and to entrain salt from below (Dickson et al., 1996). Additional strong cooling then achieved a vigorous deep convection event in early 1972. Afterwards convection occurs again every year until the end of the time series. The upper layer returns to more saline and dense conditions but remains cool, in a state clearly different from Phase 1. Possibly this is a consequence of the deep convection chimney being farther away from the weather ship now; the larger winter gap between upper and lower layer salinity (compared to Phase 1, see Fig. 2.1 b) suggests this. The deep layer jumps back to a colder and less saline state, and a further cooling and

freshening trend sets in, albeit with little effect on density.

In summary, the OWS Bravo data show a transition from a state of annual convection to stable stratification and back to convection. These transitions are at the core of our understanding of the stability of deep convection that is presented in this work.

2.3 The two-box model of deep convection

The analysis of the OWS Bravo data motivates a two-box model of a potentially convective water column in the open Labrador Sea. Welander (1982) had the basic idea fur such a model, and it was further developed by Rahmstorf (2001) and Kuhlbrodt et al. (2001). The model is designed to study the interplay of the slowly changing fluxes caused by advection and eddy diffusion with the short and vigorous convective mixing events. As sketched in Fig. 2.4, the model consists of two stacked boxes. A shallow upper box (index 1) represents the surface layer of the ocean that is well-mixed throughout the year. A large deep box (index 2) represents the waters below the seasonal thermocline that are involved in deep convective mixing. The box depths are assumed to be constant, their ratio being termed h^*. In this way, effects like variable convection depth or mixed layer deepening are not included. The advantage of keeping the model so simple is the possibility of solving the model equations analytically (section 2.5).

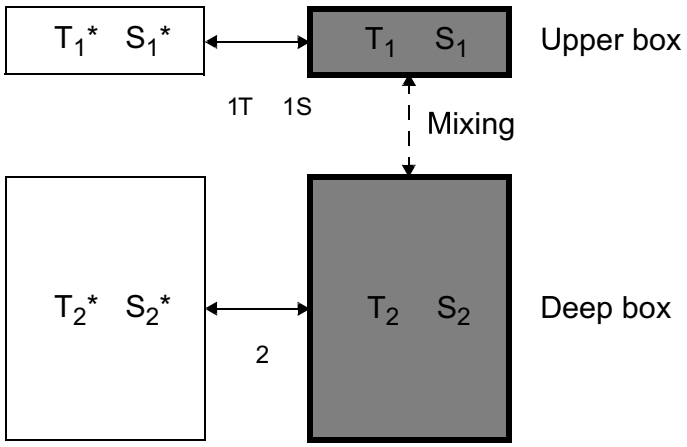

Figure 2.4: *Sketch of the 2TS box model with 2 active boxes and four variables (temperature T and salinity S in each box). Solid arrows denote continuous processes like eddy mixing and advection, while the broken arrow represents the non-continuous convective adjustment.*

The variables of the two-box model are the temperatures T_1, T_2 and the salinities S_1, S_2 in both boxes – hence it is called the 2TS model. These four variables are relaxed towards prescribed relaxation temperatures and salinities T_1^*, S_1^*, T_2^*, S_2^*. This relaxation represents exchange processes with the surrounding waters. Since the water column has open boundaries everywhere, it cannot be considered as a closed system, and the model equations cannot be derived from conservation laws.

The relaxation includes various processes for the single variables. The upper box temperature T_1 is coupled strongly to the atmosphere through surface heat fluxes. The upper box salinity S_1 is partly determined by the freshwater flux (evaporation minus precipitation) balance at the sea surface, without a feedback to the atmosphere. This fundamental difference is accounted for by using two different time scales τ_{1T} and τ_{1S} for the restoring. In addition, this restoring includes heat and salt exchanges with the surrounding waters, mostly by lateral eddy mixing. The deep box temperature T_2 and salinity S_2 are assumed to be determined by the eddy transfer fluxes at depth, which motivates a common restoring time scale τ_2. Rahmstorf (2001) further refined the model by adding a seasonal cycle with amplitude A_T to the upper box relaxation temperature T_1^* to include seasonality and achieve a short winter convection period instead of year-round convection, in accordance with the observations. To facilitate comparison with observations, and given the pronounced seasonal cycle in upper layer salinity in the data, Kuhlbrodt et al. (2001) added a seasonal cycle (with amplitude A_S and zero mean) to the upper box salinity forcing as well. A phase shift ψ between the temperature and salinity cycles is introduced as an adjustable parameter. Finally, keeping the dimensions of the variables clarifies the physical meaning of the parameters. Time is in units of 1 year. The full equation set now reads:

$$\frac{dT_1}{dt} = \frac{1}{h^*\tau_c(\Delta\rho)}(T_2 - T_1) + \frac{1}{\tau_{1T}}\left(T_1^* - A_T\cos(2\pi t) - T_1\right) \tag{2.1}$$

$$\frac{dS_1}{dt} = \frac{1}{h^*\tau_c(\Delta\rho)}(S_2 - S_1) + \frac{1}{\tau_{1S}}\left(S_1^* + A_S\cos(2\pi t + \psi) - S_1\right) \tag{2.2}$$

$$\frac{dT_2}{dt} = \frac{1}{\tau_c(\Delta\rho)}(T_1 - T_2) + \frac{1}{\tau_2}(T_2^* - T_2) \tag{2.3}$$

$$\frac{dS_2}{dt} = \frac{1}{\tau_c(\Delta\rho)}(S_1 - S_2) + \frac{1}{\tau_2}(S_2^* - S_2)\,. \tag{2.4}$$

In each of the equations the first term on the right hand side represents convective vertical mixing, and the second term the horizontal and surface heat/salt flux. The crucial difference between the comparatively slow advective and diffusive fluxes on the one hand and the quick and short convection events on the other hand is reflected in the fact that the four relaxation temperatures and salinities, the three relaxation time-scales, the two amplitudes and the phase shift are model parameters (see Table 2.2 for a detailed

list), but that the time scale of vertical convective mixing τ_c is a strongly nonlinear function of the model variables through the density difference

$$\Delta\rho = \rho_1 - \rho_2 = -\alpha(T_1 - T_2) + \beta(S_1 - S_2)\,. \tag{2.5}$$

Here, α and β are the thermal and haline expansion coefficients of the linearized equation of state of seawater (Gill, 1982). Carrying out the linearization around 4°C and 35 psu yields $\alpha = 0.105$ kg m^{-3} K^{-1} and $\beta = 0.796$ kg m^{-3} psu^{-1}.

For stable stratification ($\Delta\rho \leq 0$) the vertical mixing is very weak, so τ_c has a large value. In the case of unstable stratification ($\Delta\rho > 0$) convection starts, i.e., vigorous vertical mixing with a time scale τ_c of the order of days. Since this value of τ_c is much smaller than the other involved time scales, we assume $\tau_c \to 0$ and use the common parameterization for deep convection, known as convective adjustment (Rahmstorf, 1993; Klinger et al., 1996). The water column is checked at each time step for hydrostatic stability (we used $dt = 2$ days). Nothing is done in case of stable stratification, but any occurring instability is instantaneously removed by complete mixing. Thus the numerical integration scheme has two parts.

1. Integrate forward (2.1) to (2.4) one time step without the vertical mixing terms. If we start at time i, this gives preliminary values $\hat{T}_1^{i+1}$, $\hat{S}_1^{i+1}$, $\hat{T}_2^{i+1}$, $\hat{S}_2^{i+1}$ for the variables at time $i+1$.

2. Apply the convective adjustment scheme to obtain the final values T_1^{i+1}, S_1^{i+1}, T_2^{i+1}, S_2^{i+1} of the variables. If $\Delta\rho \leq 0$, the final values are identical to the preliminary ones; if $\Delta\rho > 0$, the two columns are mixed:

$$\begin{aligned} T_1^{i+1} = T_2^{i+1} &= h^* \hat{T}_1^{i+1} + (1-h^*)\hat{T}_2^{i+1} \quad \text{and} \\ S_1^{i+1} = S_2^{i+1} &= h^* \hat{S}_1^{i+1} + (1-h^*)\hat{S}_2^{i+1}\,. \end{aligned} \tag{2.6}$$

The convective adjustment (CA) even enhances the nonlinearity $\tau_c(\Delta\rho)$, resulting in a step function. Although this strong nonlinearity leads to numerical problems in OGCMs (Cessi, 1994; Molemaker and Dijkstra, 2000; Lind et al., 2002; Titz, 2002), CA is widely used because more sophisticated parameterizations are often too expensive in terms of computing time. Here, CA is chosen since it keeps the model simple, allowing analytical solutions. Experiments with a continuous function $\tau_c(\Delta\rho)$ (e.g. a hyperbolic tangent) yielded very similar results. However, to obtain realistic model behaviour, $\tau_c(\Delta\rho)$ must be very steep close to $\Delta\rho = 0$. This renders the equation system rather stiff, bringing up the need for special algorithms with variable time step for the numerical integration. Those algorithms have two consequences that are negative in the context of our study: first, computing time increases considerably, cancelling out an advantage of the simple model; second, a variable time step is hard to reconcile with algorithms for the solution of stochastic differential equations, which we need in chapters 3 and 4.

2.4 Fitting the two-box model to the OWS Bravo data

We adjust the model parameters (see Table 2.2 for a complete list) to find the best fit of the model to the OWS Bravo data, and use a least squares fit procedure for this purpose. We define a cost function K as the sum of the quadratic distances of each monthly averaged model variable time series T_1, S_1, T_2, S_2 (weighted by the thermal and haline expansion coefficients to have a common density unit) to the observed values. The optimal parameter set minimizes K. As discussed above, the OWS Bravo data show two different phases with convection, but the model can have only one convecting solution with the same forcing. Hence we restrict our analysis to Phases 1 and 2 of the OWS Bravo data. (An analysis based on Phases 2 and 3 gives similar results.) The 10-dimensional parameter space is spanned by a 10-dimensional matrix. For every possible parameter combination out of this matrix the cost function K is computed from a model run with convecting state initial conditions. The onset of the GSA in the model is achieved by adding an anomalous salt flux of -0.8 psu/yr to the upper box for a period of three months in spring 1968 to mimic the arrival of an advected freshwater anomaly. The crucial idea here is that we aim to find one single parameter set that yields a realistic model behaviour in both states (convecting and nonconvecting) with the same forcing; the prescribed salt flux anomaly provides a short anomaly which induces a state transition in the model.

It turns out that K and its derivatives with respect to the parameters are smooth functions and behave in a physically understandable manner. However, both physical intuition and objective analysis show that the ten free model parameters are under-determined by the fit: the problem is ill-posed. This is because the OWS Bravo data contain a steady convecting state (Phase 1), but not a steady nonconvecting state. Phase 2 of the OWS Bravo data displays the initial trends after cessation of convection but does not reveal which equilibrium values the three variables S_1, T_2 and S_2 would eventually reach in the nonconvecting state. Only the trend in T_1 stops in 1970, so that T_1^* can be determined. The impact of the missing nonconvecting state on the parameter determination can be clarified through the model equations. Take (2.2) for the upper box salinity. We neglect the seasonal cycle here because it has no impact on the long-term trend. If no convection occurs, (2.2) reduces to

$$\frac{dS_1}{dt} = \frac{1}{\tau_{1S}} \left(S_1^* - S_1\right) . \tag{2.7}$$

When dS_1/dt and an initial S_1 are known from the data, then on the right-hand side of (2.7) for any arbitrary choice of S_1^* a corresponding value of τ_{1S} can be found to fulfil the equation. Thus one of the two parameters in (2.7) is free. The situation is similar for T_2^* and S_2^*, and since (2.3) and (2.4) are

Parameter	Value	Uncertainty	Description
T_1^*	4.4°C	4.0 – 4.6°C	Upper box restoring temperature
S_1^*	33.5 psu	#	Upper box restoring salinity
T_2^*	4.1°C	3.9 – 4.3°C	Deep box restoring temperature
S_2^*	34.97 psu	#	Deep box restoring salinity
τ_{1T}	5 months	3 – 9 months	Restoring time scale of upper box temperature
τ_{1S}	8 years	6 – 11 years	Restoring time scale of upper box salinity
τ_2	20 years	14 – 28 years	Restoring time scale of deep box
A_T	6.4°C	5.0 – 7.8°C	Amplitude of seasonal cycle added to T_1^*
A_S	4.5 psu	3 – 6 psu	Amplitude of seasonal cycle added to S_1^*
ψ	0.6 months	−0.5 – 1.5 months	Phase shift of the seasonal cycles
h^*	1/36	#	Ratio of box depths

Table 2.2: *The model parameters with the values determined from fitting the 2TS model to the OWS Bravo data. The set of these parameter values is called the "optimal parameter set". The uncertainty range is spanned by all parameter sets for which the cost function value exceeds the minimum by less than 10%. The parameters marked with # were not determined through the cost function, but directly from observational data.*

coupled through the common time scale τ_2, a second degree of freedom arises here. In short, the least squares fit procedure constrains the ten-dimensional parameters space to a two-dimensional subspace. This subspace is clearly seen when attempting to minimize the cost function.

Two further constraints are thus needed to close the problem, i.e. to obtain a global minimum in K. One could assume arbitrary values of τ_{1S} and τ_2; we opt for making assumptions about S_1^* and either T_2^* or S_2^*. This option is equivalent to assuming the equilibrium mean values of the three variables S_1, T_2, S_2 in the nonconvecting state. The upper box salinity is expected to lie between the 34.7 psu of the convecting Labrador Sea and the approximately 32 psu of the North Pacific at the same latitude. From the mean salinity distribution in the Labrador Sea (Levitus, 1982) a value of S_1^* = 33.5 psu seems plausible. For the deep box parameters T_2^* and S_2^*, the average values for the waters in the North Atlantic in the latitude band of the Labrador Sea (but excluding the Labrador Sea itself) and between 200 m and 2000 m depth are $T_2^* = 4.3°\mathrm{C}$ and $S_2^* = 34.97$ psu. These values are almost equivalent to assuming Irminger Sea conditions for the nonconvecting Labrador Sea. They yield a deep box time scale of $\tau_2 = 20$ yr.

Finally, we applied one further assumption, namely to restrict the length of the winter convection event t_{wce} to less than 20 days, as low cost function values were in some cases reached also with excessively long convection periods. All the major conclusions of the present work are insensitive to the somewhat arbitrary assumptions described in this paragraph and hold for a wide range of S_1^*, T_2^*, S_2^* and maximum t_{wce}. The constraints from the OWS Bravo data are sufficient to determine the stability properties discussed below.

The fitting procedure now arrives at a global minimum of the cost function K and is repeated with parameter matrices of higher resolution (in parameter space) to localize the global minimum more exactly. The optimal parameter set thus determined is shown in Table 2.2. The value for the eleventh parameter h^* is a result of our analysis in section 2.2 and not of the fitting procedure. Changes in h^* (by assuming an upper box depth of 100 m, say) lead to quantitatively slightly different results. To measure the parameter uncertainty we determined all parameter combinations for which the cost function remains within 10% of its minimum value. This defines an uncertainty range for each parameter, save the fixed ones.

In comparison with other studies on exchange rates in the Labrador Sea (Khatiwala and Visbeck, 2000) the values for the time scales are rather large. Yet, firstly this is in agreement with the small trends in the deep layer during the GSA (see Table 2.1), and secondly our parameters apply for both the convecting and nonconvecting state, in contrast to the study of Khatiwala and Visbeck (2000) that deals with the state of frequent deep convection only. We could not use more sophisticated algorithms for estimating the parameters from the observational data because these algorithms cannot

cope with the strongly nonlinear behaviour of the convective adjustment. Algorithms that estimate the parameters for stochastic differential equations (e.g. the unscented Kalman filter) fail for the same reason (H. Voss, pers. comm.). Hence for this study we content ourselves with the optimal parameter set.

A comparison of a model run using the optimal parameter set with the OWS Bravo data (Fig. 2.5) shows that many relevant features of the data are captured by the model. This includes the two upper box seasonal cycles during Phase 1 leading to a winter convection event each year. The minimum difference between S_1 and S_2 is very small in the model time series because the model mixes the two boxes completely, whereas for a number of reasons the complete mixing is not visible in the observational data. In this cyclostationary state of the model, without stochastic forcing, there is no interannual variability in any variable. Phase 2 starts with a negative salt flux anomaly in the model run. Convection ceases, and the model reproduces the observed trends of all four variables: the upper box cools and freshens strongly while the deep box warms and becomes more saline. (The trends in the deep box are hardly visible on the scale of Fig. 2.5.) In accordance with the observational data, the model deep box is not an infinite reservoir of heat and salt (like in former versions of the box model, e.g. Welander (1982) or Lenderink and Haarsma (1994)), but receives diffusive fluxes from the neighbouring waters. The end of Phase 2, marked by the beginning salinification of the upper layer in the OWS Bravo data, is achieved in the model by a cold and saline anomaly in the upper box forcing. Only a strong anomaly in surface forcing is capable of turning convection on again. It is clearly seen from the trends in Fig. 2.5 that the longer convection is off, the lighter the surface layer becomes and the stronger an anomaly must be to restart convection. This positive salinity feedback is studied in greater detail in the next chapter. After convection is started again the model returns to its previous convecting state. The different Phase 3 state of the OWS Bravo data cannot be captured by the model.

2.5 Stability of Labrador Sea deep convection

The stable states of the deterministic 2TS model under varying parameters are now explored. For any given set of parameters, a model state is called stable if small perturbations can excite only finite excursions in a neighbourhood of the state, and if after such a perturbation the stable state is eventually reached again. (Section 4.2 provides a more rigorous definition of stability.) The states are cyclostationary due to the presence of the seasonal cycles. The numerical stability analysis performed here, together with the analytical results achieved by Rahmstorf (2001), gives a complete picture of the model's stability properties. However, we omit here an oscillatory

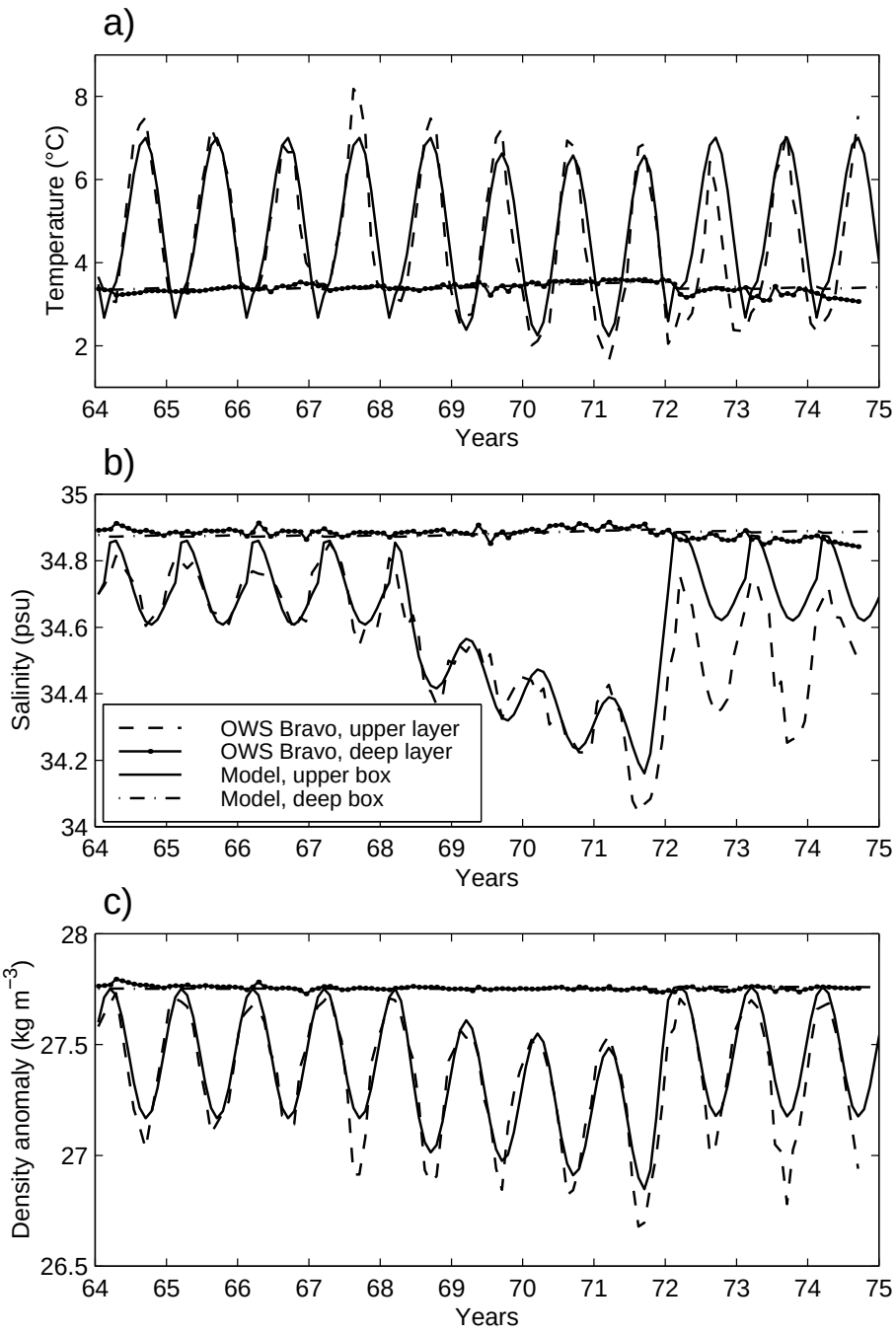

Figure 2.5: *Comparison of model output (solid, upper box; dash-dotted, deep box) and observed time series (dashed, upper layer; dotted, deep layer) for temperatures (a), salinities (b), and densities (c). The model was run with the optimal parameter set. All graphs are in monthly means. The GSA, also called Phase 2 in the text, was started by adding a negative salt flux anomaly in the upper box during April to June 1968. The GSA was stopped by a positive salt flux anomaly during October to December 1971 accompanied by a cold anomaly in the upper box temperature forcing.*

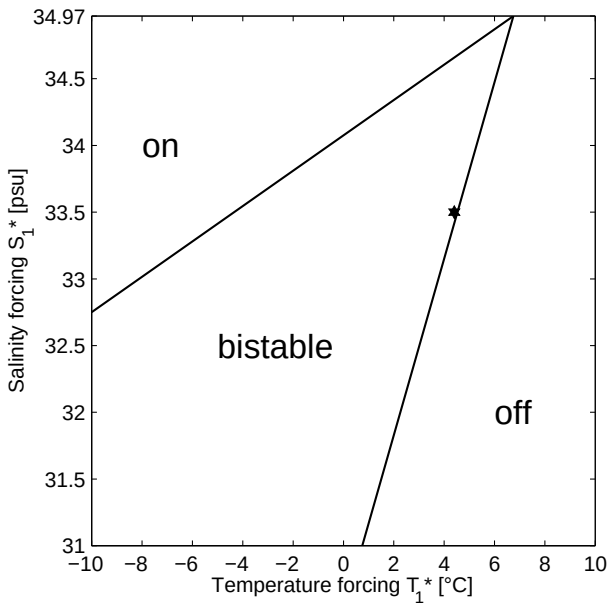

Figure 2.6: *Stable model states depending on the upper box buoyancy forcing parameters. Convection is either "on" or "off". These two stable states overlap in a bistable domain. The asterisk denotes the position of the optimal parameter set in the parameter space. The stable states were determined for varying only T_1^* and/or S_1^*, keeping the other parameters from the optimal set constant. For $S_1^* > 34.97$ psu the upper box would be forced to a higher salinity than the deep box; this case does not occur in high-latitude deep convection like in the North Atlantic. For other parameter sets within the uncertainty range or with different values of S_1^* or S_2^* the slopes of the domain boundaries change, but the marginal position of the parameter set itself is a robust feature. The inclusion of sea ice effects would affect the diagram for low T_1^* values.*

state which appears for parameter sets that do not represent the conditions in the North Atlantic. Rather, we are interested in stability changes in response to the upper box buoyancy forcing, given by T_1^* and S_1^*. Changing other parameters will lead to similar pictures. The parameter space section along these two axes (Fig. 2.6) shows the stable model states: the convecting state and the nonconvecting state. They overlap in a bistable domain. The domain boundaries are given in a good approximation by the analytical expressions for the necessary conditions for the stability of the nonconvecting state:

$$\alpha(T_1^* - A_T - T_2^*) > \beta(S_1^* - A_S - S_2^*) \tag{2.8}$$

and for the stability of the convecting state:

$$\alpha(T_1^* - A_T - T_2^*) < \beta(S_1^* - A_S - S_2^*)\frac{\tau_{1T} + \tau_2 h^*}{\tau_{1S} + \tau_2 h^*}\,. \tag{2.9}$$

These expressions are easily derived from the model equations (2.1) to (2.4). For the nonconvecting stable state, the density difference (2.5) for the equilibrium values of the four variables has to be negative. For the convecting stable state, the density difference between the mixed water column and the nonconvecting deep box (both in equilibrium) has to be positive in order to maintain convection. Since we introduced seasonal cycles such that convection now occurs around the temperature minimum and the salinity maximum in the upper box, the respective amplitudes are subtracted and added. (See also (9) and (12) in Rahmstorf (2001).) The shape of Fig. 2.6 is in agreement with earlier studies (Lenderink and Haarsma, 1994). We can thus conclude that the presence of the seasonal cycles does not change the basic stability properties of Welander's (1982) box model.

The position of the model Labrador Sea – represented by the optimal parameter set – in the parameter space section is marked by the asterisk in Fig. 2.6. It is in the bistable domain: both states, convecting and nonconvecting, are steady states of the model under the given conditions. A sufficiently large anomaly can switch convection on or off. Moreover, the model is located very close to the domain where only the nonconvecting state is stable. Changing the buoyancy forcing by a few tenths of a degree or a few tenths of a psu will lead to the convecting state becoming unstable. In other words, there are two possible ways for suppressing convection. In the first way, convection is temporarily switched off by an anomaly, but can be restarted later by an opposite anomaly, while the average properties of the buoyancy forcing (i.e. the model parameters) do not change. This is the GSA case depicted in Fig. 2.5, and this may be the case with a 50 yr-long spell of convection in the Labrador Sea being switched "on" among centuries without convection, which was found by Tett et al. (1997) in a coupled GCM. In the second way, under a slowly changing buoyancy forcing (e.g., T_1^* or S_1^*) the convecting state eventually becomes unstable and convection stops. This scenario could apply to the global warming GCM run of Wood et al. (1999), in which Labrador Sea convection stops early in the 21^{st} century. The simplified 1S model (see chapter 4) will provide a means to understand these two ways of stopping convection in detail. The first way is then a hopping between two potential wells, and the second way is equivalent to a disappearance of one well. The role of the anomalies triggering state transitions in both cases is examined in detail in chapter 3.

We checked systematically how the stability diagram changes for different choices of S_1^*, S_2^* (or T_2^*) and maximum t_{wce}, as well as for parameter sets in the uncertainty range defined by a 10% change in the cost function. Eqs. (2.8) and (2.9) show that those parameter changes affect the width of

the bistable domain: for instance, for a larger difference $(S_1^* - S_2^*)$ the bistable domain (in terms of T_1^*) is wider. However, the distance of the model solution from the border of the bistability domain varies only by some tenths of a degree on the T_1^* axis and similarly small amounts on the S_1^* axis. This is a consequence from the OWS Bravo data which show that a freshwater anomaly equivalent to -0.2 psu induced a transition form the convecting to the nonconvecting state. The precarious position of the Labrador Sea, in the bistable domain but close to the nonconvecting domain, is therefore a robust feature of the fitted model.

2.6 Conclusions

A new picture of the deterministic stability of open ocean deep convection has been developed in this chapter. This new picture is based on the analysis of observational data. Depending on the external forcing, the water column may have two stable states, one with regular deep convection events and the other without deep convection. Additional anomalies in the forcing trigger transitions between these two stable states. Convection can be switched off in two ways: either temporarily, by an anomaly, or definitively, if continuous parameter changes drive the model out of the bistable domain.

The analysis tool is a two-box ("2TS") model of the water column in a potentially convective part of the ocean. The model is rather simple (only four variables), but comprehensive enough to reproduce observed time series in a satisfying way. This is an advantage over earlier studies that used simpler versions of the convection box model and dealt with model output only. The 2TS model opens up the possibility to estimate the model parameters from observations.

The observed data are from the Labrador Sea. They show three distinct phases between 1964 and 1974, marked by presence or absence of deep convection events in winter. The four-year phase without deep convection is characterized by a freshening and cooling in the upper layer (a "Great Salinity Anomaly").

With the model parameters obtained from a least squares fit procedure, the position of the Labrador Sea in a stability diagram can be determined. For a certain region in the parameter space the model has two stable states, with convection being either "on" or "off" each winter. Using the Labrador Sea parameters the model is located in this bistable domain, meaning that anomalies in the forcing are capable of triggering jumps from the convecting to the nonconvecting state and back.

The model shows that lasting anomalies similar to the Great Salinity Anomaly (1968–72) can be triggered by short-term anomalies in the surface conditions suppressing convection in one winter. This mechanism of transitions between two stable states can explain the basic properties of the time

series from OWS Bravo, and is thus a part of the answer to the first guiding question (GQ1). Compared with other studies this is a new understanding. According to Lilly et al. (1999), the atmospheric heat flux forcing is predominant in triggering the very convection events, and the long-term freshwater advection sets the background stratification. They conclude that it was persistent weak atmospheric forcing that led to the absent deep convection in the years 1969–1971. Dickson et al. (1996) suggest that deep convection was suppressed in two complementary ways: wind anomalies over Greenland enhanced freshwater advection, while other wind anomalies over the Labrador Sea itself reduced the the cyclonic wind stress and the heat fluxes.

In those pictures the potentially convective water column merely reacts passively to the external forcing. Following our main hypothesis, the external forcing kicks off the local internal dynamics of the water column. A short-term perturbation switches convection off, and the subsequent evolution is governed by the local positive salinity feedback (Fig. 1.2) until another perturbation switches convection on again. At least for the perturbation that terminated the GSA in late 1971 this understanding can be substantiated by observations (Dickson et al., 1996; Mertens, 2000): deep convection was switched on again by local weather forcing, not by internal ocean processes. Moreover, the positive salinity feedback is apparently necessary to close the freshwater budget of the upper Labrador Sea (Houghton and Visbeck, 2002). High-resolution ocean models also show this interaction of short anomalies with the local bistability (sect. 1.3; (Lenderink and Haarsma, 1994, 1996; Rahmstorf, 1995c)). All this evidence renders our main hypothesis fairly plausible.

It turns out that the position of the model Labrador Sea in the bistable domain is very close to the border to the monostable domain where the convecting state is unstable. This holds true irrespective of the parameter assumptions about the stable nonconvecting state. This position is precarious: changing the ocean's surface forcing by about one degree towards warmer conditions leads to the convecting state becoming unstable. If convection is switched off in this way, the convection shutdown is abrupt and definitive. An external perturbation can cause at most a transient convection event, but no jump back to regular deep convection is possible. This is another type of sensitivity of deep convection and thus another part of the answer to GQ1. Such a shutdown of Labrador Sea convection did occur in a global warming scenario computed with a coupled GCM (Wood et al., 1999).

Model simulations and observational data show that the longer a nonconvecting phase lasts, the harder it is to interrupt it. This effect is due to the positive salinity feedback. It means that if deep convection in the Labrador Sea once stops in reaction on global warming, it takes an extraordinarily strong anomaly to make it start again.

The results of Wood et al. (1999) suggest that in a complex, high-resolution ocean model deep convection in the Labrador Sea is similarly as easy to interrupt as in the box model used here. One of the main differences between simple box models and complex GCMs is the variability that is present in the latter. Therefore, the influence of variability on the deterministic stability properties discussed in this chapter needs to be studied. The next two chapters show how such external variability excites internal variability in the box model, and how it changes the stability properties.

Chapter 3

Variability of deep convection excited by stochastic forcing

3.1 Overview

This chapter explores the role of stochastic forcing for the variability of deep convection. The presented approach follows the concept of Stochastic Climate Models and the theory of Large Deviations, which are outlined in section 3.2. The stochastic forcing is meant to represent the short-term weather variability. With such a forcing term included, the 2TS model is called the stochastic climate model of deep convection (section 3.3). The parameters for the stochastic forcing term are estimated from an observed heat flux time series. The main effect of the stochastic forcing are the jumps between the convecting and the nonconvecting model state. The model output shows events that resemble the Great Salinity Anomaly. This corroborates the hypothesis that short anomalies, here provided by the stochastic forcing, can trigger interannual convective variability. Deep convection events occur even if the convecting state is unstable (section 3.4). This observation is a first step towards the concept of stochastic stability, in which the mean residence times in the two model regimes play a crucial role (section 3.5). A pronounced asymmetry between the two regimes appears for the distributions of the residence times. It turns out that long-term interruptions of deep convection can be part of the natural variability, and that this is a consequence of the positive salinity feedback (section 3.6). The concluding section 3.7 summarizes the differences between the deterministic and the stochastic model and highlights how the results help to understand the variability of deep convection in the North Atlantic.

3.2 Theory of stochastic climate models

"A characteristic feature of climatic records is their pronounced variability." With this statement Hasselmann (1976) starts his seminal paper about the theory of stochastic climate models. As the application of his methods to study the variability of deep convection events is the foundation of the next two chapters of our work, we give here an outline of the theory of stochastic climate models. This includes mentioning briefly the theory of Large Deviations (Freidlin and Wentzell, 1998) which extends Hasselmann's ideas in a way that is essential for the present work. (More details about Freidlin and Wentzell's theory are found in chapter 4.) We also refer to Arnold (2001) who re-expressed Hasselmann's and Freidlin and Wentzell's ideas highlighting the main mathematical steps.

It is remarkable that the observed variability of most climatic variables is distributed over a very broad range of time scales: from the seconds of local turbulence phenomena, over the strong diurnal cycle, the synoptic variability of the weather on the scale of days, the strong seasonal cycle, decadal, centennial and millenial variability up to the ice age cycles of 100,000 years. In other words, the spectral variance of climatic variables extends over a frequency range that spans many orders of magnitude.

Often it is observed that the variance is stronger at lower frequencies. The variance spectrum is then said to be "red". This red spectrum is reproduced by the highly complex, deterministic general circulation models (GCMs), but any simpler *deterministic* model stays far from explaining this special shape of the spectrum. However, the most simple *stochastic* differential equation does reproduce this red spectrum. This fact was one of Hasselmann's main motivations to introduce stochasticity into climate modelling.

A fundamental assumption of stochastic climate models is the *time scale separation.* Suppose we have a GCM, given by a system of differential equations

$$\dot{z} = h(z)\,. \tag{3.1}$$

It is now assumed that the components of z come in two well separated kinds: fast, "weather" variables y with a dominant time scale of days, and slow, "climate" variables x with typical time scales of months, years and longer. Typical fast variables are atmospheric variables like air temperature or sea level pressure; typical slow components of the climate system are the ocean, ice shields or the biosphere. The question whether the assumption of time scale separation can be reliably founded on observations is still under discussion and exceeds the scope of the present work. We make this assumption nevertheless, bearing in mind that Hasselmann's ideas have had some success, starting from an immediate sequel paper (Frankignoul and Hasselmann, 1977) to very recent work (e.g. Ganopol-

ski and Rahmstorf (2002)). Since the two time scales are clearly different, (3.1) is now written as a coupled system of differential equations

$$\begin{aligned} \dot{x} &= f(x,y) \quad \text{“climate” variables} \\ \dot{y} &= \frac{1}{\varepsilon} g(x,y) \quad \text{“weather” variables}\,. \end{aligned} \tag{3.2}$$

The parameter ε is assumed to be small. For the typical time scales, we then have $\tau_x \approx 1 \gg \tau_y \approx \varepsilon$.

Provided that we do not want to solve the complete equation system (3.1) or merely are not capable of doing so due to lack of computer power, we can now take advantage of the time scale separation. If we are interested only in the evolution of the mean of the climate variables – their first moments –, we can make use of the *Method of Averaging.* Most Statistical Dynamical Models (SDMs) are obtained in this way. Given a differential equation

$$\dot{x} = f(x, \xi_t^\varepsilon) \tag{3.3}$$

with a fast forcing function ξ_t^ε, one can derive an averaged right-hand side by defining

$$F(x) = \langle f(x, \xi_t^\varepsilon) \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x, \xi_t^\varepsilon)\, dt\,. \tag{3.4}$$

The averaging is carried out for fixed x, i.e. we consider time scales $\tau < 1$. In most cases one assumes ergodicity, meaning that the time average in (3.4) may be replaced by the mean over an ensemble of realizations. The averaged differential equation, expressed in the variable $u = \langle x \rangle$, is now given by F:

$$\dot{u} = F(u)\,. \tag{3.5}$$

With ε controlling the smallness of ξ_t^ε, one can show that, for diminishing ε, the solution x of the full equation (3.3) converges to the solution u of the averaged equation (3.5) on a finite time interval $[0, T]$. Since the fast variability has been averaged out to obtain the averaged equation, the so-called SDMs of the form (3.5) are indeed rather deterministic than statistical. Note that (3.3) does not contain a feedback from the slow variables to the fast ones. The procedure outlined here can also be carried out for the slow part of (3.2) with the full coupling between the fast and the slow variables. Yet, in this case it is not always ensured that x converges to u in the small noise limit.

If we want to have information about the variance of the climate variables – their second moments – as well, we can use the *Central Limit Theorem.* Consider the deviation of the full solution from the averaged solution

$$x^\varepsilon - u\,, \tag{3.6}$$

where we have written x^ε to denote the dependence of the full solution x on ε. Under the condition that the fast process ξ_t^ε has a short decorrelation

time and does not depend on x, the Central Limit Theorem now states that in the limit $\varepsilon \to 0$ and on a finite time interval $[0, T]$, the process

$$\frac{1}{\sqrt{\varepsilon}}(x^\varepsilon - u) \tag{3.7}$$

has a normal Gaussian distribution. In other words, the error made by the Method of Averaging is asymptotically normally distributed.

The Central Limit Theorem is the reason why Gaussian noise processes are so widely used in climate modelling: if the details of those fast processes are not known or not resolved one can still rely on the fact that their distribution is, in the limit, Gaussian. Since we now know how the first and second moments of the solution of (3.3) behave, we can write

$$\dot{x} \stackrel{D}{=} F(u) + \sigma\zeta_t \tag{3.8}$$

with $F(u)$ describing the average evolution, ζ_t a normally distributed Gaussian process with short decorrelation time, and σ a noise intensity. The equality $\stackrel{D}{=}$ is an approximate equality in distributions. If we consider the evolution of (3.8) on short time scales ($\tau < 1$), the averaged solution may be assumed to be constant, such that $F(u) = x$. The solution of (3.8), interpreted as a proper equation, and driven by a white noise process ζ_t, is then a random walk, with its variance growing without bounds.

On time scales of $\tau \approx 1$, effects like negative feedbacks and dissipation will come into play, influencing the behaviour of the averaged solution. The simplest way to include those is to assume $F(x) = -kx$, such that

$$\dot{x} = -kx + \sigma\zeta_t\,. \tag{3.9}$$

The solution of this simple stochastic differential equation has a clearly defined, finite variance, and the variance spectrum is indeed red. Thus, (3.9) is the basic form of a Stochastic Climate Model. The slow components of the climate provide the damping feedback, determining the average evolution. They "integrate" the fast atmospheric variability that is approximated as a Gaussian random process. This very picture is Hasselmann's idea of Stochastic Climate Models, and the Method of Averaging and the Central Limit Theorem are their theoretical underpinning.

We note in passing that the picture of a slow system component driven by fast motion has been developed about one hundred years ago to describe the motion of a pollen grain in a liquid. The description of this Brownian motion was historically the physically motivated way into the theory of distributions of random variables and into stochastic differential equations (e.g. Einstein (1905)).

The Method of Averaging and the Central Limit Theorem are valid only on finite time intervals. But what happens if we consider time intervals

growing with $1/\varepsilon$, if we are interested in the system's evolution on time scales much larger than $\tau_x \approx 1$? On such large time scales, arbitrary large deviations of the fast variables (or noise processes) from the averaged motion might occur. In case that there is one stable steady state of the averaged solution, the non-averaged solution of (3.9) will then leave any bounded neighbourhood of this stable state, notwithstanding the fact that the variance is bounded. In the case that there is more than one stable steady state, the full solution can hop between the different attractors. This effect adds a new quality to the system: where the deterministic solution can always reach only one attractor, the stochastic system can escape to other attractors or regimes. In fact, it will do so after a typical time scale, the *mean escape time* $\langle t_e \rangle$. The theory of *Large Deviations* (Freidlin and Wentzell, 1998) derives an analytical expression for $\langle t_e \rangle$. Suppose we have a simple system, driven by Gaussian white noise, where a potential difference ΔU has to be overcome to leave the neighbourhood of interest. One then has

$$\lim_{\sigma \to 0} \sigma^2 \ln \langle t_e \rangle = \Delta U \,. \tag{3.10}$$

Thus, the mean escape time depends exponentially on the potential difference and the square of the noise intensity σ^2:

$$\langle t_e \rangle \approx e^{\Delta U / \sigma^2}. \tag{3.11}$$

Freidlin and Wentzell (1998) generalized this result for a much wider class of dynamical systems by introducing a *quasipotential*. The quasipotential is defined as the minimal work required to reach the boundary of the neighbourhood E_1 from the stable state x^*. This definition does not assume the dynamical system to have a potential. The only requirement is that there is some way to compute the work needed to move the system along a given path in phase space. Equation (3.11) is applied in this sense in the present chapter. In the next chapter, (3.11) together with a proper potential U is at the heart of the concepts of stochastic stability (section 4.2).

The Method of Averaging and the Central Limit Theorem are the fundamentals of linear stochastic climate models which have only one stable state. The Large Deviations theory is needed in addition for nonlinear models with more than one stable state, like the box models used in this study.

3.3 A stochastic climate model of deep convection

A particular physical motivation for introducing variability in the 2TS box model has become clear in chapter 2: anomalies in the surface buoyancy forcing are capable of switching convection on and off. Hasselmann's theory suggests to include the variability as a stochastic forcing term. Observations (Marshall and Schott, 1999; Lilly et al., 1999) suggest that heat flux

anomalies (due to weather activity) and freshwater flux anomalies (by lateral mixing) act together to trigger or suppress convection. Yet, the heat fluxes clearly prevail in their contribution to the overall buoyancy forcing (Sathiyamoorthy and Moore, 2002). Hence, we focus on heat flux variability as the primary variability component. The characteristic time scale of synoptic cyclones is a few days. Applying the concept of a stochastic climate model, we parameterize synoptic-scale variability by a stochastic term added to the surface heat flux forcing. A parameter estimation, given in detail below, suggests the stochastic term to consist of red noise ξ_t with a decorrelation time of about six days times a standard deviation σ. The model equation for the upper box temperature (2.1) is thus extended to

$$\frac{dT_1}{dt} = \frac{1}{h^*\tau_c(\Delta\rho)}(T_2 - T_1) + \frac{1}{\tau_{1T}}\left(T_1^* - A_T\cos(2\pi t) + \sigma\xi_t - T_1\right). \quad (3.12)$$

Equation (3.12) together with the other three model equations (2.2) – (2.4) and the convective adjustment (2.6) yields the stochastic 2TS box model of deep convection.

Note the special role of the convective adjustment (CA). The model has two distinct regimes, with the two boxes being either uncoupled (convection "off") or fully mixed at each time step (convection "on"). As long as the model trajectory is in the nonconvecting regime, the Hasselmann ansatz of time scale separation is justified. The smallest dynamical time scale of the model ($\tau_{1T} = 5$ months) is still much larger than the time scale of the stochastic forcing term ($\tau_D = 6$ days). Yet, due to the CA we expect the model to switch to the convecting regime from time to time. The short time scale of the convective mixing, expressed through the instantaneous CA, is not in full agreement with the assumed time scale separation. The physics of the deep convection process however motivates to have two fast processes in the model: the convective mixing and the stochastic forcing. We deal with this issue again in the framework of the 1S model presented in chapter 4.

In the stochastic 2TS model, two classical conceptual models – for deep convection (Welander, 1982) and for high-frequency atmospheric forcing of the ocean (Hasselmann, 1976) – are combined to give a simple model that we call the *stochastic climate model of deep convection.* For its numerical integration we applied a semi-implicit Milstein scheme following Kloeden and Platen (1992). Its theoretical convergence is twice as good as with an ordinary Euler scheme, which is relevant regarding the discontinuity arising from the CA.

Although applying Hasselmann's concepts suggests to work with variance spectra, their use is not feasible here. The model time series can become non-stationary due to the CA. This leads to spurious kinks in the variance spectra, making their physical interpretation difficult. Therefore we use and develop other methods to analyze the variability seen in the model.

In general, the fundamental physical differences between deterministic and stochastic differential equations would require a completely new parameter estimation for the stochastic climate model of deep convection. Yet, this requirement is not easy to fulfil. Algorithms that estimate parameters for stochastic differential equations are more complex than the simple least squares method we used in section 2.4. If applied to our model, they fail due to the strong nonlinearity posed by the convective adjustment (H. Voss, pers. comm.). For this reason only the parameters for the stochastic term are estimated here, and the other parameters (table 2.2) are taken over.

The time series from OWS Bravo (section 2.2) has a too small sampling rate for estimating the parameters of the weather forcing. Instead, we analyzed a 52-year long time series of daily net surface heat flux at the OWS Bravo site. This time series was extracted from the NCEP database (Kalnay et al., 1996). The aim is an estimate for the standard deviation and the decorrelation length of the noise term in (3.12). Assuming that the heat flux time series $Q(t)$ can be decomposed into an average flux Q_0, a seasonal cycle with amplitude A_Q, and a noise term ξ_t with standard deviation σ_Q,

$$Q(t) = Q_0 + A_Q \cos(2\pi t) + \sigma_Q \xi_t \,, \tag{3.13}$$

it turns out that, for averaging intervals of one to a few days, the autocorrelation function of the noise process ξ_t falls off to zero only after the first few lags. In other words, ξ_t can be modelled by an AR(1) process

$$\xi_t = \alpha_1 \xi_{t-1} + \zeta_t \,, \tag{3.14}$$

where the value ξ_t at time t is determined by the value ξ_{t-1} at the previous time step times the autocorrelation at lag 1, α_1, plus a random value ζ_t from a Gaussian white noise process. With values for α_1 estimated from the NCEP time series, the decorrelation time as defined in von Storch and Zwiers (1999),

$$\tau_D = \frac{1+\alpha_1}{1-\alpha_1} \,, \tag{3.15}$$

lies between five and seven days, depending on the averaging interval of the time series. This decorrelation time is just the typical time scale for synoptic activity. Using this red noise forcing in the model (instead of pure white noise) is a more realistic parameterization of the high-frequency heat flux variability and renders the model results more robust against changes in the time step of the numerical integration scheme. The dependence of the estimates τ_D on the averaging interval is a fundamental feature (Wunsch, 1999). Here we use the time step that is motivated by the typical time scale for deep convection events, which is two days.

Apart from the decorrelation length, the second parameter we need to estimate for the noise term in eq. (3.12) is the standard deviation σ. Using

the heat flux time series in eq. (3.13) to force the temperature of an ocean surface layer we write

$$c_p \rho_0 h_1 \frac{dT}{dt} = Q_0 + A_Q \cos(2\pi t) + \sigma_Q \xi_t - \lambda T \,, \tag{3.16}$$

with the specific heat capacity $c_p = 3990$ J kg^{-1} K^{-1}, a reference density $\rho_0 = 1028$ kg m^{-3}, the surface layer depth $h_1 = 50$ m, and a restoring constant λ. Setting $c_0 = c_p \rho_0 h_1$, this reads

$$\frac{dT}{dt} = \frac{Q_0}{c_0} + \frac{A_Q}{c_0} \cos(2\pi t) + \frac{\sigma_Q}{c_0} \xi_t - \frac{\lambda}{c_0} T \,. \tag{3.17}$$

Comparison with (3.12) shows how the standard deviation σ of the stochastic forcing in (3.12) is related with the standard deviation σ_0 of the heat flux time series:

$$\sigma = \sigma_Q \frac{\tau_{1T}}{c_0} \,. \tag{3.18}$$

With a dependence on the averaging interval again, we find $\sigma_Q = 120$–140 W m^{-2}, which translates into $\sigma = 8$–9°C. This is a rather high value, larger than the average flux Q_0 – and still the seasonal cycle of the standard deviation itself, reaching its maximum in winter, has not been accounted for here. Sathiyamoorthy and Moore (2002), in their analysis of the buoyancy flux at OWS Bravo derived from weather ship data, obtain a similar result. The dominant role of the synoptic-scale variability of the heat flux in the Labrador Sea is highlighted here again. For simplicity we have added noise only to the surface temperature forcing. To obtain a realistic variability in the whole surface buoyancy flux pronouncedly higher values of σ have to be assumed since the variability of surface freshwater fluxes and of eddy-induced transports of heat and salt have not been considered here. A range of $\sigma = 15$–20°C seems therefore plausible.

3.4 Stochastic forcing and state transitions

Thirty years from a model run with stochastic forcing are displayed in Fig. 3.1. Several times, convection is interrupted for a few years, indicated by the large minimum density difference $\Delta\rho$ between the upper and the deep box (Fig. 3.1c). The upper box temperature T_1 (Fig. 3.1a) is the only variable directly influenced by the noise, so it shows the strongest variability. Apart from the convective mixing induced by that variability the other three variables evolve in an unperturbed way. Similarly to the GSA in the OWS Bravo data (Fig. 2.1), in the nonconvecting years the upper box tends to freshen (Fig. 3.1b) and cool (Fig. 3.1a), until a cold anomaly restarts convection again. In contrast to the observed GSA the nonconvecting phase in this model run is kicked off by a warm anomaly in the upper box, not a

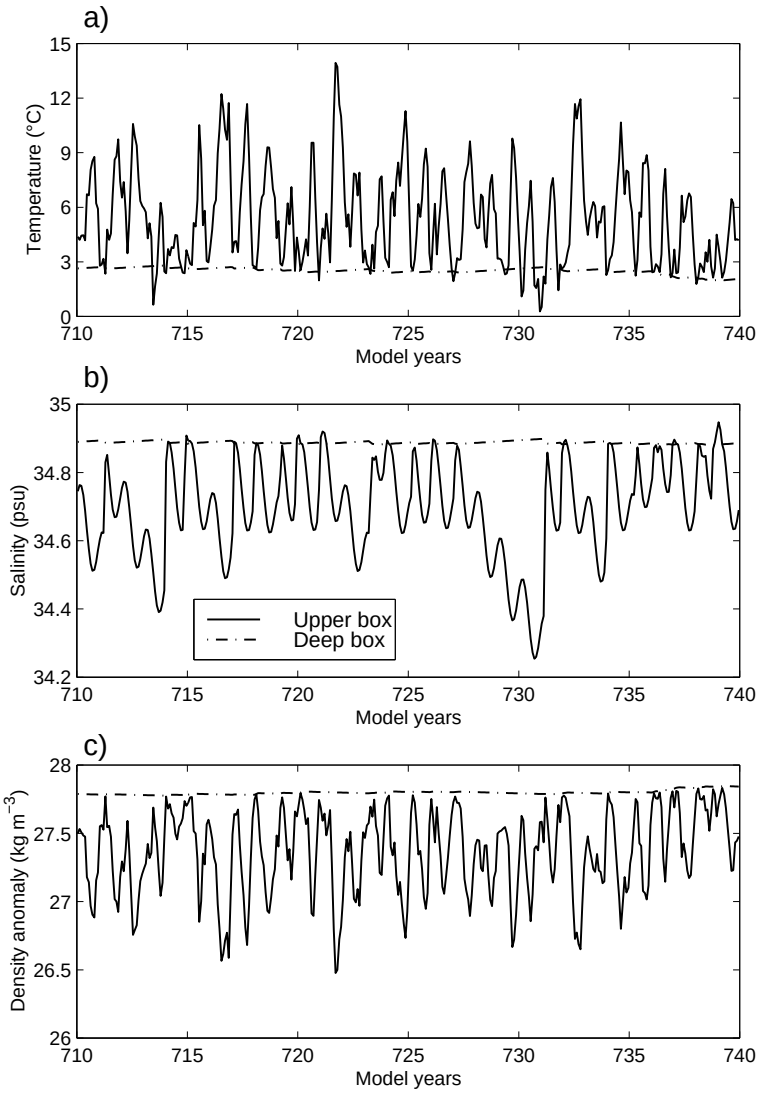

Figure 3.1: *Time series of monthly means from the stochastically forced model: temperature (a), salinity (b), and density (c) of the upper box (solid) and the deep box (dash-dotted). The model was run with the optimal parameter set and a standard deviation* $\sigma = 18°C$ *of the stochastic forcing. The thirty model years shown include several interruptions of the convecting state. The difference to the observed GSA (cf. Fig. 2.1) is that the positive buoyancy anomaly needed to stop convection is achieved here by local heat fluxes rather than lateral freshwater fluxes.*

freshwater anomaly. This is due to the fact that the stochastic variability appears in T_1 only, not in S_1, so by construction only temperature anomalies can appear. However, through the Hasselmann mechanism the upper box integrates the weather noise to (intra-)seasonal temperature anomalies (because $\tau_{1T} = 5$ months). Thus, integrated synoptic heat flux anomalies and advected intra-seasonal freshwater anomalies have the same impact on $\Delta\rho$.

The key point here is that a comparatively short anomaly (lasting for a couple of months) triggers the state transition, and that the long term trends evolving afterwards are due to the internal dynamics until the next anomaly triggers the next state transition. We hypothesize that this picture of a bistable water column holds for the Labrador Sea. This implies that the falling surface salinity from 1968 to late 1971 does not result from anomalous freshwater input during this whole time, nor does the return to convecting conditions in 1972 result from an end to the anomalous forcing. Rather, the falling salinity only requires a short trigger anomaly (which could even have been thermal instead of freshwater) that prevents convection in 1968. Convection then cannot recover by itself but requires another substantial trigger event. The longer convection has been off the larger the trigger needs to be. Had the winter of 1972 not been such a harsh one, subsequent winters would have needed to be even colder to restart Labrador Sea convection.

This hypothesis is consistent with the conclusion of Dickson et al. (1996), who analyzed the 1972 convection onset in the Labrador Sea in detail in the observed data. They conclude that the jump-like rise of the upper layer salinity is explicable only by anomalously strong wind forcing that mixes saline intermediate waters into the mixed layer; advection or mixing processes cannot lead to such strong changes. In other words, observational data show that the termination of the GSA in the Labrador Sea was achieved by anomalous weather conditions at the ocean's surface, not by its internal dynamics.

The presence of noise leads to a qualitative change in the model's stability behaviour: the sharp domain boundaries depicted in Fig. 2.6 are replaced by more gradual changes in the frequency of the occurrence of convection. As a measure we use n_c, the fraction of years with convection out of all years in a model run. Fig. 3.2 shows how n_c depends on the standard deviation σ of the noise. Using the optimal parameter set with the convecting state as initial condition and increasing σ (Fig. 3.2a), n_c drops close to zero for weak noise. This reflects the marginal position of the optimal parameter set in parameter space (Fig. 2.6). Any small perturbation shifts the model into the nonconvecting state, but the small perturbations are not able to induce a jump back to convection. For $\sigma > 12°\mathrm{C}$, the convecting state is reached in some cases. The fraction of convecting years rises quickly and asymptotically reaches a value of $n_c \approx 0.75$. When the noise is strong, it tends to override the deterministic stability properties. This feature becomes clearer

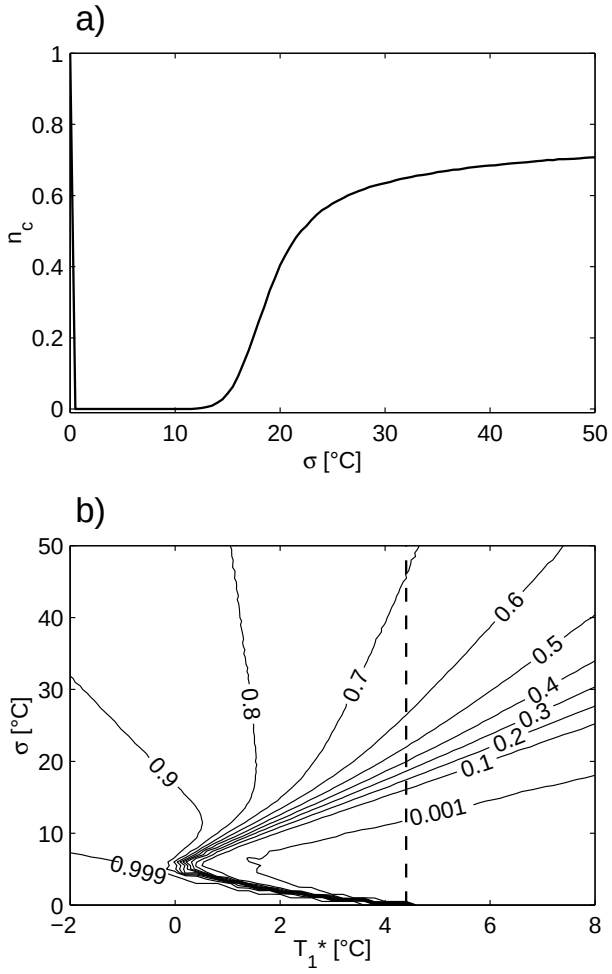

Figure 3.2: *(a) Dependence of the fraction n_c of convection years in a long model run (10^6 yr) on the noise level σ. The model was run with the unchanged optimal parameter set and convecting state initial conditions. The deterministic case without stochastic forcing is at $\sigma = 0$. (b) Contour plot of n_c as a function of σ and additionally of the upper box temperature forcing T_1^*. The dashed line denotes the position of the graph shown in (a). With a nonconvecting initial condition, $n_c = 0$ for $\sigma = 0$ for all values of $T_1^* > -4.4°C$ (i.e. in the bistable domain, see Fig. 2.6). The lower part of (b) would change accordingly.*

when considering changes in the surface buoyancy forcing T_1^* in addition. The contour plot in Fig. 3.2b displays how n_c depends on σ and T_1^*. For $\sigma = 0$, Fig. 3.2b corresponds to the deterministic (not stochastically forced) parameter space section (Fig. 2.6) with only the nonconvecting state being stable for $T_1^* > 4.5°\mathrm{C}$. For low noise and large T_1^* there is a large, wedge-shaped domain where almost no convection events occur. The shape of this domain can be understood when thinking of the convecting state becoming less and less stable for larger T_1^*. Then, for larger T_1^* a smaller amount of noise is needed to trigger a jump into the nonconvecting state. Since the nonconvecting state becomes more stable for larger T_1^*, a larger amount of noise is necessary to trigger jumps back into the convecting state. Fig. 3.2b also shows that beyond this wedge-shaped domain the noise is capable of keeping the model in the convecting state for more than half of the time even when this state is unstable in the deterministic case. The exact extent of this domain depends on the respective parameter set: for instance, a lower S_1^* leads to a less dense upper box in the nonconvecting state, so convection is harder to trigger, and the domain becomes wider. The concept of a state becoming "less" or "more" stable is put on a theoretical foundation in chapter 4.

In section 3.3 it was shown that for the Labrador Sea conditions in the box model σ is likely to be near or larger than 15°C. This means that the model is located in a domain where n_c is sensitive to changes in the surface forcing. There are two ways of making convection occur less often: either by decreasing the variability σ or by increasing the surface temperature T_1^* (or equivalently decreasing S_1^*). But convection can still occur even when the convecting state is unstable in the deterministic case. If an increase in T_1^* is taken as a crude representation of a global warming scenario, then these results suggest that the frequency of Labrador Sea convection could decrease substantially due to a future warming (and/or freshening) unless variability increases strongly at the same time.

3.5 Residence times and stability

Apart from the mere frequency of convection it is also of interest how long a continuous sequence of convecting or nonconvecting years lasts. The length of such a sequence is dubbed *residence time*. The mean residence times in the two model states can be used to estimate how the stability of the model states depends on the model parameters.

Consider a particle in an ideal double-well potential. If the particle is initially in one well, then added noise will rattle it. To leave the initial potential well and jump to the other one, the particle has to overcome a potential difference ΔU. If the noise is Gaussian distributed, eventually one perturbation, occurring after time t_e, is large enough for the particle to hop

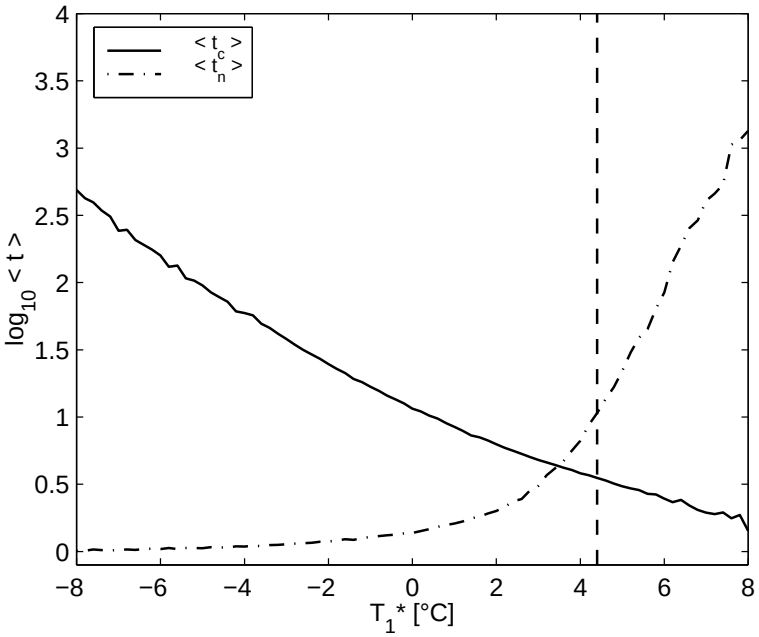

Figure 3.3: *Mean residence times in the convecting state ($\langle t_c \rangle$, solid) and in the nonconvecting state ($\langle t_n \rangle$, dash-dotted) in dependence on the mean surface temperature forcing T_1^* for constant standard deviation ($\sigma = 18°C$) of the stochastic surface temperature forcing. The dashed line indicates the position of the optimal parameter set: the convecting state lasts 3.5 yr and the nonconvecting state 11.2 yr, on average.*

into the other well. In this ideal case (treated in chapter 4) we can use the relation (3.10) between the particle's mean escape time $\langle t_e \rangle$ from one well, the potential difference ΔU and the intensity σ of the added noise,

$$\lim_{\sigma \to 0} \sigma^2 \ln \langle t_e \rangle = \Delta U \,, \tag{3.19}$$

to characterize the stability of the states. The two stable states of the box model (with four variables) cannot be expressed as minima of a (one-dimensional) potential. However, there is a way to use (3.19) for our purposes. The potential U can be interpreted as a *quasipotential* (sect. 3.2). Then, the potential difference ΔU is the necessary perturbation strength for a state transition. The larger ΔU is, the more stable is the state. Instead of the mean escape times of eq. (3.10), we use the mean residence times $\langle t_c \rangle$ and $\langle t_n \rangle$ in the convecting and the nonconvecting state as estimated from long model runs. We do this for leaving σ constant and varying T_1^* only. With the help of (3.19) we are then able to give a quantitative estimate of the relative stability of the two states as a function of the surface forcing. Fig. 3.3 shows how the logarithms of the residence times change

with varying T_1^*. For low values of T_1^*, the convecting state is clearly the more stable one. Conversely, for warm surface forcing the stability of the nonconvecting state increases strongly.

There are two caveats when interpreting the residence times in this way. First, (3.19) makes sense only if there is a *deterministically* stable state from which the stochastic system may escape. In Fig. 3.3 both stable states coexist only for $-4.4°\mathrm{C} < T_1^* < 4.5°\mathrm{C}$. Thus, their stochastic stability in terms of the mean residence time can be compared only on this interval. Second, the noise has to be small relative to the potential difference. As the potential difference approaches zero close to the respective domain border, the above T_1^* interval has to be chosen even smaller. Note, however, the contrast between the sharp stability domain borders of the deterministic model (Fig. 2.6) and the smoothly shaped graphs of $\ln\langle t_r\rangle$ as a function of T_1^* in the stochastic case. Since T_1^* from the optimal parameter set is close to the interval boundaries, we cannot make a quantitative comparison of the stabilities; we just observe that the nonconvecting state is clearly more stable than the convecting one. This gives a quantitative understanding of Fig. 3.2a: weak noise can provide the anomalies to jump into the nonconvecting state, but anomalies twice as large, necessary for the jump back, occur only extremely rarely.

From Fig. 3.3 we see that the mean residence times for the optimal parameter set and a standard deviation of $\sigma = 18°\mathrm{C}$ in the stochastic forcing are $\langle t_c\rangle = 3.5$ yr and $\langle t_n\rangle = 11$ yr. Thus, the average time for the model to jump from one state to the other and back is about 15 yr. In other words, the typical time scale for the variability is in the decadal range. This is clearly different from the synoptic time scale of the stochastic surface forcing. The effect of the weather noise is here to excite intraseasonal variability in the mixed layer, which in turn triggers interannual to decadal variability. The deep ocean, being isolated from the atmosphere nearly at all times, "sees" the synoptic variability through the "window" of deep convection events – but responds to this forcing with its own typical decadal time scale. In this way deep convection is an example of time scale interactions in the climate system.

3.6 Distributions of residence times

The analysis of residence times can be carried one step further by extending our view from their *means* $\langle t_c\rangle$ and $\langle t_n\rangle$ to their *distributions* $p_s(t_c)$ and $p_s(t_n)$, which are equivalent to the stationary *probability density functions* (pdfs). This draws a statistically more complete picture of the variability. The different shapes of the two pdfs in Fig. 3.4 stand out. The residence times in the convecting state (Fig. 3.4a) are distributed following a straight line, with $t_c > 20$ yr occurring only

rarely during a 100,000 yr model run. In comparison, $p_s(t_n)$ has a bent shape with high probability density for very short residence times and some occurrences of t_n exceeding 20 yr. These features are obscured when considering the mean value only: while the mean residence time is

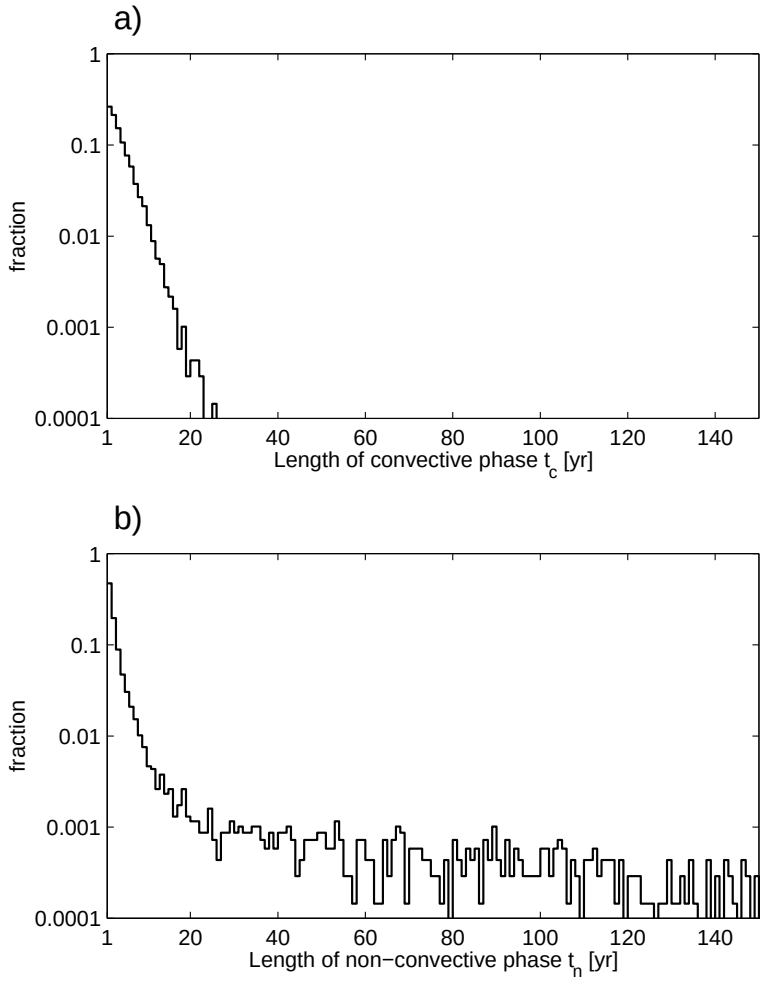

Figure 3.4: *Distribution of residence times in the convecting state (a) and the nonconvecting state (b). The frequency of every single residence time is given as a fraction of the total number of residences during a* 10^5 *yr run. Hence, the distributions are approximate probability density functions. The fraction axis is logarithmically scaled; the small bar in panel (a) at* $t_c = 25$ *yr corresponds to one single occurrence. The model was run with the optimal parameter set and a standard deviation of* $\sigma = 18°C$*, yielding a fraction of convecting years* $n_c = 0.26$*. The tail of the distribution in (b) was cut arbitrarily; the maximum* t_n *is 526 yr.*

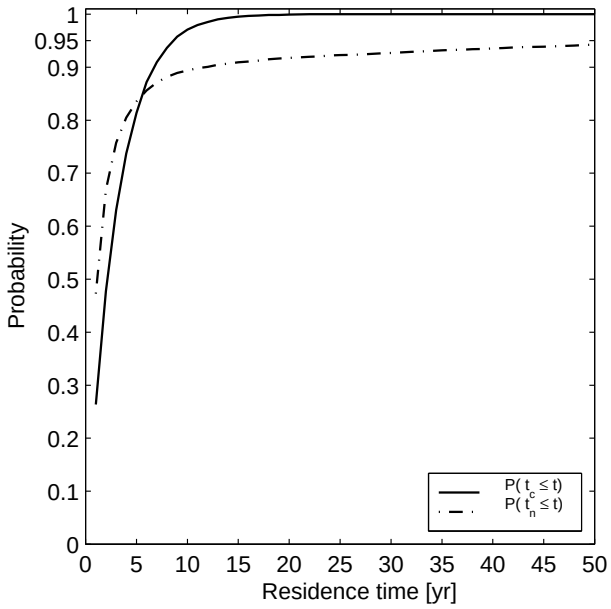

Figure 3.5: *Probability distribution of residence times in the convecting state (t_c, solid) and the nonconvecting state (t_n, dash-dotted) from the same model run as in Fig. 3.4. For any time t, the probability distribution gives the probability of the residence time being smaller than or equal to t.*

$\langle t_n \rangle = 11$ yr, we learn from $p_s(t_n)$ that the time series will contain many cases of only a few years without convection, but also some occasions where convection is interrupted for more than 100 years. With the *probability distributions* of the residence times (Fig. 3.5) this can be quantified. For instance, there is a 10% probability for the nonconvecting state to last longer than 13 yr, but the convecting state will do so only with a probability of 1.5%. Hence, the observation of two decades without deep convection, like in the years 1982–2001 in the Greenland Sea (Rhein (1996); Visbeck and Rhein (2000); J. Holfort (pers. comm.)), is not necessarily a sign of a global climatic trend but could be within the natural variability properties of a convecting water column.

The difference between the two distributions in Fig. 3.4 can be understood qualitatively in the framework of "runs" introduced by von Storch and Zwiers (1999). A "run" is defined as the time that a stochastic process spends uninterruptedly on one side of its mean value. Von Storch and Zwiers analyzed AR(1) processes with varying autocorrelation coefficient α_1, and found that for $\alpha_1 = 0$ (white noise) the run length pdf decreases exponentially. In a logarithmic plot this pdf of the run lengths, or residence times, appears as a straight line like in Fig. 3.4a. For red noise ($\alpha_1 > 0$), long residence times are more probable at the expense of intermediate times, which

gives the bent shape of the graph in Fig. 3.4b. In other words, the linear shape of $p(t_c)$ in Fig. 3.4a means that the probability for a convection stop is equal in all years, whereas the exponential shape of $p(t_n)$ in Fig. 3.4b stems from the diminishing probability of leaving the nonconvecting state with increasing residence time t_n. This feature again reflects the positive salinity feedback.

3.7 Conclusions

The influence of surface forcing anomalies on the variability of deep convection has been explored in this chapter. The theory of stochastic climate models justifies to model those anomalies by a stochastic forcing term. Frequent jumps between the convecting and the nonconvecting state occur. The variability arising from these jumps is characterized through the residence times in both states. The convection frequency depends smoothly on the forcing parameters, giving a new view on the stability of deep convection that contrasts with the abrupt stability changes in the non-stochastic model of chapter 2. The positive salinity feedback leads to a non-negligible probability for long nonconvecting phases.

The stochastic 2TS model joins Welander's (1982) deterministic concept of the stability of convection with Hasselmann's stochastic concept of climatic variability. This stochastic climate model of deep convection is based on our main hypothesis: that short anomalies trigger switches between the two states of the model. These anomalies are now provided by the stochastic forcing. However, it is not the small fluctuations around a deterministic stable state that are of interest here – the interesting events are the large fluctuations which carry the system close to the other deterministically stable state. Where Hasselmann's theory does not focus on the influence of those large fluctuations, they are the core topic of the Large Deviations theory of Freidlin and Wentzell (1998). Their theory is applied to the stochastic 2TS model here. Thus, joining the concepts of Welander and Hasselmann leads to a new picture of the stability and variability of deep convection that goes beyond both those concepts.

Chapter 2 showed that there are two ways to switch between the two model states: temporarily, caused by anomalies, or definitively, by changing a model parameter (e.g., T_1^* or S_1^*). In the stochastic model there are anomalies triggering convection switches all the time, much like in the real ocean. The physically interesting observable is now the frequency of deep convection events. Now, if one studies the effect of changing parameters like T_1^* or S_1^* in the stochastic model, one observes that the convection frequency is a continuous function of the forcing parameters.

This is the core result of this chapter, and the main answer to GQ2: in the stochastic model, the frequency of convection event depends smoothly

on the heat and freshwater forcing of the upper box. An abrupt loss of stability like in the deterministic 2TS model does not occur. Even when the convecting state is unstable in the deterministic case the variability excites frequent convection events.

Frequent convection switches between phases with and without deep convection are also seen in CGCMs (P. Wu, pers. comm.), and the local convective bistability appears as well in OGCMs (Rahmstorf, 1995a; Lenderink and Haarsma, 1996). These examples suggest that convective variability is indeed driven by our proposed mechanism. However, this needs to be corroborated by closely analyzing those models with our box model, and by studying other time series from deep convection sites. This could include adding more vertical layers to the box model, as well as coupling several boxes horizontally (as done by Titz (2002) and Lind et al. (2002)).

The frequency of convection events depends on the noise strength in a highly nonlinear way. There are two plateaus for the frequency value: weak noise triggers hardly any jumps, and with strong noise the jumps become very frequent and occur every few years. In between is a small range of noise strengths where the jump frequency steeply rises. For the plausible parameter range the position of the Labrador Sea is in the region of this steep rise, which reflects its sensitivity to changes in surface forcing.

With the convection frequency depending both on the forcing parameters and the noise strength, these two effects may reinforce or cancel out each other. For instance, a mean freshening, as may be expected in the near future due to global warming, could be compensated in its effect on convection frequency by a stronger surface forcing variability. This result from the stochastic approach answers GQ1 in a different way than with the deterministic approach of chapter 2.

Density anomalies in the upper box can be caused either by the heat flux forcing or by the freshwater flux forcing. In the stochastic 2TS model, the stochastic heat flux forcing with its time scale of days is integrated by the upper box to give intraseasonal temperature anomalies. From observations it is seen that salinity anomalies of the same intraseasonal time scale cause density anomalies as well. The impact of both types of anomalies on the vertical density gradient is similar. Since the focus is here on the statistics of deep convection events, it is sufficient to have one source of anomalies only. The impact of stochastic freshwater anomalies is studied in the 1S model of chapter 4.

The intraseasonal anomalies in the upper box lead to interannual to decadal phases during which convection is permanently "on" or "off". Thus, the 2TS box model shows a way how weather variability, being integrated in the surface layer, can excite decadal variability in the deep ocean. An example of this mechanism being at work is given by Weisse et al. (1994). In their OGCM, they have just the freshwater anomaly advection mechanism we propose as equivalent to the temperature anomalies. The deep water

formation area in their model is south of Greenland, not in the Labrador Sea. The Hasselmann mechanism generates surface salinity anomalies in the Labrador Sea. Since no deep convection occurs there, these anomalies are advected to the deep water formation site, where they trigger decadal variability of the ocean circulation.

The probability distribution of the residence times in the nonconvecting state shows that there is a small, but not negligible probability for the nonconvecting state lasting a decade or longer. In contrast to a deterministic understanding of the system, this means that convection may start again after a long break due to natural variability. The 19 year-long cessation of deep convection in the Greenland Sea is thus not necessarily due to a long-term climatic trend but could be part of the normal stochastic variability properties of convection. This is one more example how the stochastic model gives a new approach in explaining the variability of deep convection. In the next chapter, we will turn to its stochastic stability properties.

Chapter 4

Stochastic stability in an analytical model of deep convection

4.1 Overview

The striking differences between the deterministic model (chapter 2) and the stochastic model (chapter 3) deserve a deeper analysis. This analysis of stability and variability features is carried out in this chapter. To begin, a theoretical foundation for the stochastic stability concepts is laid in section 4.2. The tool for the theoretical analysis is a highly simplified box model of deep convection (section 4.3). Since it has only one variable, the salinity gradient, it is called the 1S model. The residence times in the two model regimes, convecting and nonconvecting, are studied in section 4.4. In order to interpret the model output in a physically meaningful way, a new probability measure is developed that combines the concepts of probability density and residence times (section 4.5). The 1S model's simplicity allows us to understand analytically how the residence times and the new probability measure depend on the model parameters. Concerning long phases of absent convection, it turns out that this dependence is not always as strong as suggested by other authors (section 4.6). The stochastic stability properties are explored in section 4.7. This includes a comparison with the deterministic stability properties, and it is shown how short, but dynamically relevant excursions to an unstable regime can be incorporated into the concept of stability. Section 4.7 has a keystone character since on the one hand the stochastic stability properties are explained analytically, and on the other hand the 1S model is shown to behave qualitatively very similar to the 2TS model of the previous chapters. This larger context is highlighted in the concluding section 4.8.

4.2 The concept of stochastic stability

We briefly discuss some results of the theory of stochastic dynamical systems. The focus is to give here a theoretical underpinning for the stability concepts developed later on in this chapter. The reader is introduced to the Fokker-Planck Equation and its application to calculate mean residence times (Gardiner, 2002), and to the theory of stochastic stability developed by Freidlin and Wentzell (1998).

4.2.1 The Fokker-Planck equation

Suppose we have a one-dimensional stochastic differential equation

$$\dot{x} = b(x,t) + \sqrt{D(x,t)}\,\zeta_t \tag{4.1}$$

driven by Gaussian white noise ζ_t. In general, b is called the drift coefficient, and D is the diffusion coefficient. The solution of this stochastic differential equation (SDE) is an individual trajectory. The Fokker-Planck equation (FPE) describes the evolution of the *probability density function* (or *pdf*) $p(x,t)$ of the diffusion process given by (4.1):

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left(b(x,t)p(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(D(x,t)p(x,t)\right)\,. \tag{4.2}$$

Now imagine that we have an over-damped particle moving in a one-dimensional potential U, driven by noise with constant intensity σ. The SDE is then

$$\dot{x} = -\frac{dU}{dx} + \sigma\zeta_t\,. \tag{4.3}$$

The corresponding FPE reads

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left(-\frac{dU}{dx}\,p(x,t)\right) + \frac{1}{2}\sigma^2\,\frac{\partial^2}{\partial x^2}p(x,t)\,. \tag{4.4}$$

Since U and σ do not depend on time, there is a solution to (4.4) that is time-independent as well. It is called the *stationary* probability density function $p_s(x)$, which is an analytical solution to (4.4):

$$p_s(x) = \mathcal{N}\exp\left[-\frac{2}{\sigma^2}U(x)\right]\,. \tag{4.5}$$

The normalization constant $\mathcal{N}$ is determined by the condition

$$\int_{-\infty}^{\infty} p_s(x)dx = 1\,. \tag{4.6}$$

The stationary probability density function is reached from almost all initial pdfs after sufficient long time.

It is clear from (4.5) that to every well in the potential $U(x)$ there is a corresponding peak in $p_s(x)$. The probability to find the system in a given interval (x_1, x_2) is

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} p_s(x)dx\,. \tag{4.7}$$

One defines the cumulative distribution function or *probability distribution*

$$P(x < x_0) = \int_{-\infty}^{x_0} p_s(x)dx\,. \tag{4.8}$$

Through (4.6) it is ensured that the probability to find the system anywhere on the x-axis equals one.

4.2.2 Measures of stochastic stability

We start with the definition for deterministic (i.e., non-stochastic) stability. Consider a dynamical system given by the differential equation

$$\dot{x} = b(x)\,. \tag{4.9}$$

Its steady states x^* are defined by

$$b(x^*) = 0\,. \tag{4.10}$$

A steady state x^* is called *stable* if any neighbourhood E_1 of x^* contains another neighbourhood E_2 such that, for $t \to \infty$, any trajectory of the dynamical system starting in E_2 does not leave E_1. The state x^* is called *asymptotically* stable if the same trajectory, in the same limit $t \to \infty$, reaches x^* itself. Put in another way, an asymptotically stable state is characterized by the fact that trajectories that start sufficiently close to it will approach it asymptotically. In the following, we say *stable* and mean *asymptotically stable*.

This stability definition includes possible transient growth. It is therefore not surprising that it still holds if we consider a dynamical system

$$\dot{x} = b(x, \psi_t) \tag{4.11}$$

with a *bounded* perturbation function ψ_t. In particular, ψ_t may be the seasonal cycle, and thus this stability definition applies to the cyclostationary states of the previous chapters.

Yet, the picture completely changes when we replace the perturbation function in (4.11) by a noise process ξ_t^σ:

$$\dot{x} = b(x, \xi_t^\sigma)\,. \tag{4.12}$$

The noise process will in general be *unbounded* in an absolute sense. One will only be able to give a boundary for the mean of the process, and one will be

able to state that there is a low, but finite probability for arbitrarily large deviations beyond this mean value. Thus, the above stability definition is not applicable any more. In the case of a Gaussian distributed noise process, the probability is in fact one that the solution of (4.12) will leave any neighbourhood E_1 of a stable state for any noise intensity σ if we only wait long enough. Using this fact, one can estimate the probability per fixed time interval for the solution to leave E_1. The reciprocal of this probability is equivalent to the mean escape time $\langle t_e \rangle$ of the solution from E_1. Suppose that we have

$$\dot{x} = b(x) + \sigma\zeta_t \tag{4.13}$$

with a Gaussian white noise process ζ_t. In addition, suppose that $b(x)$ is given by a potential U, so that the solution has to overcome a potential difference ΔU to leave E_1. It is then a classical result (Kramers, 1940) that

$$\langle t_e \rangle \approx e^{\Delta U/\sigma^2} \tag{4.14}$$

in the limit of small noise intensity σ. The mean escape time depends exponentially on the potential difference and the squared noise intensity. The higher the potential difference and the weaker the noise, the longer the solution of (4.13) will rest in its initial neighbourhood E_1. Using the 1S model which does have a potential, the mean escape time can in fact be determined exactly using the Fokker-Planck equation. For the 2TS model another approach was needed, see chapter 3.2.

Suppose we have one stable state at x^*, and we are interested in the mean escape time from the interval $(-\infty, x_0)$, where $x_0 > x^*$. When does the system leave the neighbourhood of x^*, bounded by x_0, in the positive direction? The mean escape time is

$$\langle t_e \rangle = \frac{2}{\sigma^2} \int_{x^*}^{x_0} \exp\left[\frac{2}{\sigma^2} U(x)\right] \left(\int_{-\infty}^{x} \exp\left[-\frac{2}{\sigma^2} U(x')\right] dx' \right) dx \,, \tag{4.15}$$

see Gardiner (2002). This equation is of practical use in section 4.4.

With Freidlin and Wentzell (1998) we think that (4.7) and (4.14) are useful and appropriate to characterize the stability behaviour of stochastic dynamical systems. We can measure stochastic stability with the following criteria:

1. the probability P_{x^*} to be in the neighbourhood of a stable state x^*,
2. the mean escape time $\langle t_{e,x^*} \rangle$ from this neighbourhood,
3. the potential difference $\Delta U^{(x^*)}$ needed to escape from the stable state.

These measures give us the possibility to introduce the concept of *quantitative stability*. With this concept the stabilities of two or more stable states of a stochastic dynamical system can be compared in their degree of stability (see section 4.7). Whereas the potential may often not be available, the quantities P_{x^*} or $\langle t_{e,x^*} \rangle$ can be estimated from the model output to study the stabilities (as done for the 2TS model in sect. 3.5).

Which of the three stochastic stability measures is the most convenient one? It is instructive to consider the behaviour for small noise intensity $\sigma \to 0$. The value of P_{x^*} will certainly converge to one: (4.5) tells us that the peak of p_s around the stable state becomes sharper for weaker noise, such that the values of p_s outside the considered neighbourhood decrease. With a similar argument, using (4.14), one sees that the mean escape time $\langle t_{e,x^*} \rangle$ will grow arbitrarily large for diminishing noise. The potential difference $\Delta U^{(x^*)}$ however does not depend on the noise intensity (at least in case of additive noise). This motivates choosing the deterministic quantity $\Delta U^{(x^*)}$ as a measure for stochastic stability, bearing in mind that $\Delta U^{(x^*)}$, through (4.14), determines the stochastic quantity $\langle t_{e,x^*} \rangle$. Of course, $\Delta U^{(x^*)}$ can only be used as a stability measure here because the dynamical system we are going to deal with, the 1S model, does have a potential.

Equations of the type (4.15) are valid only in the small noise limit. Yet, in the climate system the noise is not arbitrarily small, but has finite intensity. Although this puts limits to the validity of analytical expressions, it is just this finite-intensity noise that leads to the stochastic stability phenomena studied in the previous and the present chapter.

4.2.3 Time scales

It will prove useful to clarify the roles of the involved time scales in the stochastic dynamical system. Consider a stochastic dynamical system given by

$$\dot{x} = F(x) + \sigma \xi_t \tag{4.16}$$

that is driven by Gaussian noise ξ_t with a small intensity σ. Assume that $F(x)$ is a nonlinear function of x, but that $F(x)$ is well approximated linearly close to a stable state x_0:

$$\dot{x} = -k(x - x_0) + \sigma \xi_t \tag{4.17}$$

If we study how (4.16) evolves in time after a start close to x_0, we will find three relevant time scales.

1. The Gaussian noise process ξ_t is often assumed to be white noise, without any autocorrelation. Although this has advantages for the mathematical treatment, physical processes are in general observed to have a finite *decorrelation time* τ_D, a time scale on which the autocorrelation is significantly larger than zero. An example is the heat flux

time series analyzed in section 3.3. In the picture of many stochastic climate models (e.g., section 3.2), τ_D corresponds to the fast time scale of the weather variables.

2. If we wait somewhat longer, the evolution of (4.16) is governed by the feedback coefficient k in (4.17), yielding a typical *dynamical time scale* $\tau_d = 1/k$. This dynamical time scale corresponds to the slow time scale of the climate variables in the picture of stochastic climate models. It is also the time scale of the system's evolution in the deterministic limit.

3. If we observe the system still longer, then the noise will lead to excursions from the bounded neighbourhood where the linear approximation (4.17) is valid. The system might even hop between neighbourhoods of several stable states. The frequency of these excursions is given by the mean *escape time* $\langle t_e \rangle$. Roughly speaking, it depends exponentially on the deterministic term in (4.16) divided by the squared noise intensity σ^2. Therefore, in many cases the mean escape time is much larger than the dynamical time scale.

Often one focuses on the evolution on the dynamical time scale τ_d, assuming that the neighbourhood of interest is not left. This assumption may be justified, but one has to take into account that a deterministically stable state is only *metastable* under stochastic forcing: the system's trajectory is likely to be in its neighbourhood only on time scales smaller than $\langle t_e \rangle$. In the present study we are interested in the hopping events between the convecting and the nonconvecting state. Thus it is a prerequisite that the neighbourhood of either states is left frequently. The time scale considered is much larger than the mean escape time $\langle t_e \rangle$.

4.3 An analytical model of deep convection in one variable

4.3.1 Deterministic part

Our aim is to derive a minimal conceptual model of deep convection that is as simple as possible, and hence open to an analytical understanding. The starting point is the 2TS box model with four variables (temperature T and salinity S in each box) used in chapters 2 and 3. We will end up with a 1-box model of the mixed layer with only one variable (salinity), dubbed the 1S model. For the reader's ease, we repeat the equations of the 2TS model from section 2.3:

$$\frac{dT_1}{dt} = \frac{1}{h^*\tau_c(\Delta\rho)}(T_2 - T_1) + \frac{1}{\tau_{1T}}\left(T_1^* - A_T\cos(2\pi t) - T_1\right) \quad (4.18)$$

$$\frac{dS_1}{dt} = \frac{1}{h^*\tau_c(\Delta\rho)}(S_2 - S_1) + \frac{1}{\tau_{1S}}\left(S_1^* + A_S\cos(2\pi t + \psi) - S_1\right) \quad (4.19)$$

$$\frac{dT_2}{dt} = \frac{1}{\tau_c(\Delta\rho)}(T_1 - T_2) + \frac{1}{\tau_2}(T_2^* - T_2) \quad (4.20)$$

$$\frac{dS_2}{dt} = \frac{1}{\tau_c(\Delta\rho)}(S_1 - S_2) + \frac{1}{\tau_2}(S_2^* - S_2)\,. \quad (4.21)$$

The vertical exchange time scale τ_c is a function of the vertical density difference

$$\Delta\rho = \rho_1 - \rho_2 = -\alpha(T_1 - T_2) + \beta(S_1 - S_2)\,, \quad (4.22)$$

where α and β are the thermal and haline expansion coefficients of the linearized equation of state for seawater. The vertical exchange time scale is very large for $\Delta\rho \leq 0$, but for $\Delta\rho > 0$ the convective mixing starts, and τ_c is of the order of a few days.

To start the simplification, we use the fact that the variations in the deep ocean are much smaller than in the upper layer. Inspection of Fig. 2.3 tells that the difference is about one order of magnitude. This motivates setting the deep box temperature and salinity to constant values T_2^* and S_2^*. Next, the seasonal cycle is not considered. Section 2.5 showed that its presence does not change the basic stability properties. Finally, we assume that the upper box temperature is relaxed much faster than the upper box salinity, so that we can use a constant value $T_1 = T_1^*$. This is justified because in the parameters obtained from fitting the 2TS model to data from the Labrador Sea (table 2.2), the two relaxation time scales differ by a factor of about 20. With a constant T_1 there is no heat flux feedback during convection events as in the 2TS model. Therefore the length of the convection events is reduced here. For the coarse-grained criterion used later on this length is however not relevant.

All these assumptions leave us with one single equation for the upper box salinity:

$$\frac{dS_1}{dt} = \frac{1}{\tau_c(\Delta\rho)}(S_2^* - S_1) + \frac{1}{\tau_{1S}}(S_1^* - S_1)\,, \quad (4.23)$$

with the same function for τ_c as above. The vertical density gradient now depends on S_1 only:

$$\Delta\rho = \rho_1 - \rho_2 = -\alpha(T_1^* - T_2^*) + \beta(S_1 - S_2^*)\,. \quad (4.24)$$

We can rewrite the two equations (4.23) and (4.24) after a transformation of the variables, and switching from time scales to exchange coefficients:

$$\frac{dy}{dt} = -k\,y + k_S(y^* - y)\,, \text{where} \quad (4.25)$$

$$k = 0 \quad \text{for} \quad y \leq y_0$$

$$k = k_c \quad \text{for} \quad y > y_0$$

We used

$$
\begin{aligned}
y &= S_1 - S_2^* && (4.26)\\
y^* &= S_1^* - S_2^* && (4.27)\\
y_0 &= \frac{\alpha}{\beta}(T_1^* - T_2^*) && (4.28)\\
k_S &= 1/\tau_{1S} && (4.29)\\
k &= 1/\tau_c(\Delta\rho)\,, && (4.30)
\end{aligned}
$$

and $1/k_c$ of the order of a few days. For later use we define

$$K = k_c/k_S + 1\,. \tag{4.31}$$

The model is sketched in Figure 4.1. Following similar arguments, Cessi (1996a) has used the same simplifications to obtain a deep convection box model in one variable. In our 1S model, the properties of the deep ocean are fixed, and so is the upper layer temperature. The vertical salinity gradient

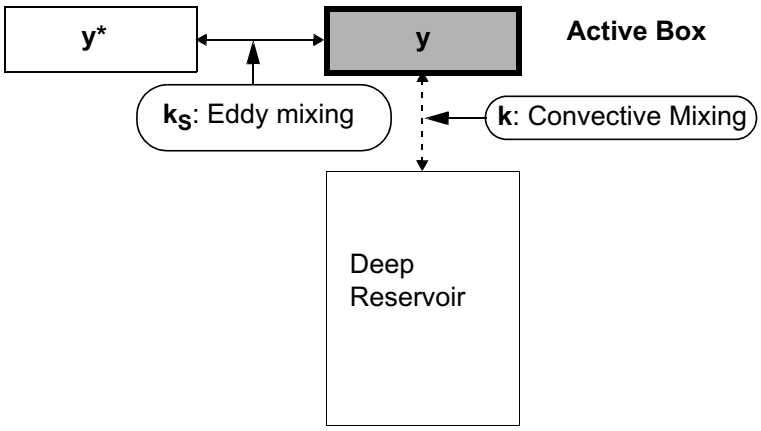

Figure 4.1: *Sketch of the 1S box model with one single active box for the surface mixed layer salinity. The only variable of the model is the vertical salinity gradient y. The deep box is considered as an infinite reservoir of water with constant salinity.*

y is restored to a reference value y^*, representing the effect of eddy mixing. Other processes acting on the upper layer are less important (Houghton and Visbeck, 2002). If the upper layer salinity becomes sufficiently high, then the vertical salinity gradient y overcomes the fixed vertical temperature gradient y_0, and convective mixing starts. Hence, y_0 plays an important role as a threshold that separates the two *regimes* of the model. In the convecting

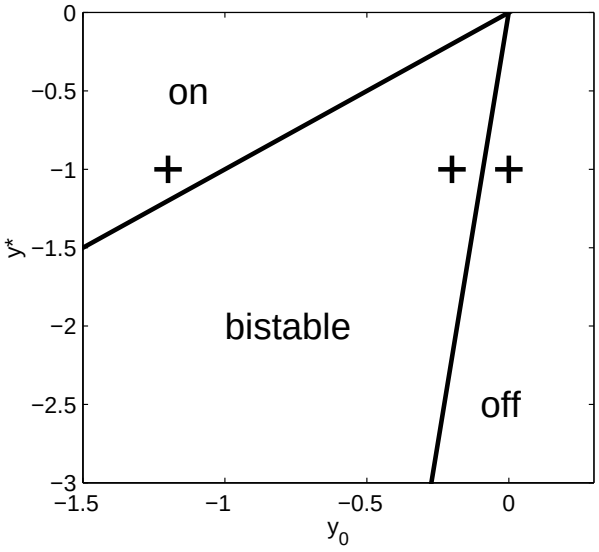

Figure 4.2: *Stability diagram of the 1S model for the parameters* $k_S = 1$ *and* $k_c = 10$*. Depending on the parameters* y^* *and* y_0 *there exist one or both stable states, with convection being "on" or "off". The lines define the borders of the respective domains. The crosses show the parameter sets used for the panels in Fig. 4.3. Note the similarity to the stability diagram of the 2TS model (Fig. 2.6).*

regime ($y > y_0$), the upper box is coupled to the deep box very strongly; in the nonconvecting regime ($y \leq y_0$) the two boxes are independent. The function $\tau_c(\Delta\rho)$ has thus been specified to be a step function having the value $1/k_c$ or ∞.

The two stable steady states $y_{st}^{(n,c)}$ immediately follow from the model equation (4.25): a nonconvecting, or "off" state at $y_{st}^{(n)} = y^*$, existing if $y^* \leq y_0$, and a convecting, or "on" state at $y_{st}^{(c)} = y^*/K$, existing if $y^* > Ky_0$. For $Ky_0 < y^* \leq y_0$ both states exist. The model is bistable in this parameter range. The stability diagram (Fig. 4.2) shows that the stability properties of the 1S model are very similar to those of the 2TS model (Fig. 2.6). With the 2TS model, the main parameters to study stability changes were the upper box restoring temperature T_1^* and the upper box restoring salinity S_1^*. In an analogous way, the fixed vertical temperature gradient y_0 and the reference vertical salinity gradient y^* are chosen here for this purpose.

A potential is readily derived from (4.25). Assuming the potential U to be zero at $y = y_0$, we define the potential function on either side of the convection threshold:

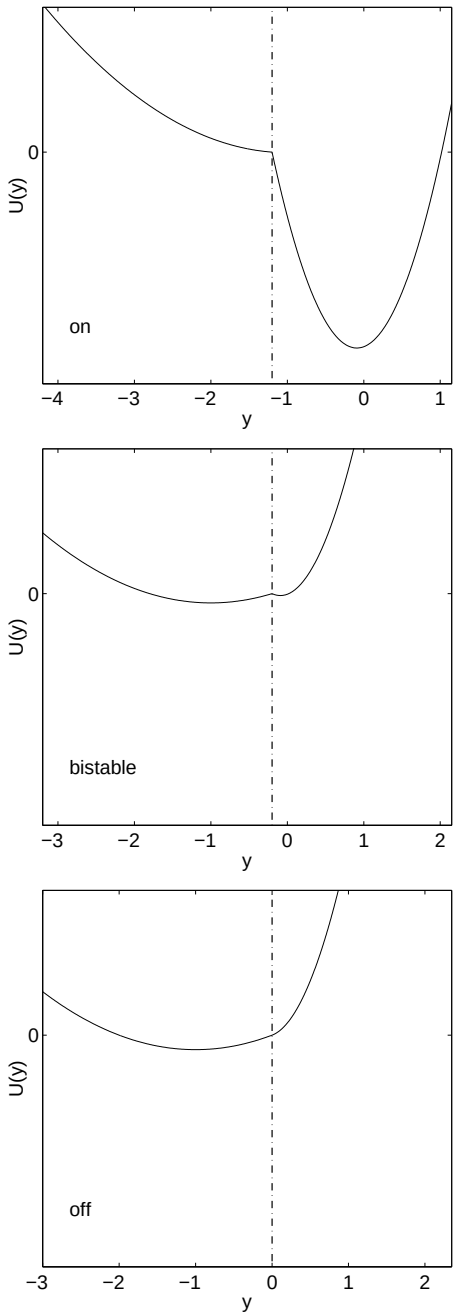

Figure 4.3: *The potential U (in arbitrary units) as a function of y as given by eqs. (4.32) and (4.33). The dash-dotted line denotes the convection threshold y_0 that separates the convecting and the nonconvecting regime. By definition, $U(y_0) = 0$. The potential is shown for three cases: convecting monostable, bistable, and nonconvecting monostable, corresponding to the crosses in Fig. 4.2. The parameters are $k_S = 1$, $k_c = 10$, $y^* = -1$, and $y_0 = -1.2; -0.2; 0.0$, respectively.*

$$U(y) = \frac{k_S}{2}\left((y-y^*)^2 - (y_0-y^*)^2\right) \quad \text{if } y \le y_0\,, \tag{4.32}$$

$$U(y) = \frac{k_S\,K}{2}\left(\left(y-\frac{y^*}{K}\right)^2 - \left(y_0-\frac{y^*}{K}\right)^2\right) \quad \text{if } y > y_0. \tag{4.33}$$

The evolution equation (4.25) can now be written as:

$$\frac{dy}{dt} = -\frac{dU}{dy}\,. \tag{4.34}$$

This potential (see Fig. 4.3) is different from a classical double-well potential in several points. First, due to the two restoring coefficients k_c and k_S the two wells are asymmetric. The strong coupling in the convecting state implies a deep potential well. Second, the convection threshold in (4.25) results in a continuous, but non-differentiable point at y_0. There is no unstable steady state at this point. Third, since the restoring coefficients are constant, the curvature of the two parabolic wells does not change with y^* or y_0; only the potential well depth relative to $U(y_0) = 0$ is altered. There are clearly defined borders for the existence of the two deterministically stable *states* (Fig. 4.2), but it is important to note that the two *regimes* always exist. Even if one of the regimes does not contain a stable state (e.g., the convecting regime in the lower panel of Fig. 4.3), this unstable regime may be accessed temporarily by the model trajectory.

4.3.2 Stochastic part

As in chapter 3, we want to study the impact of anomalies in the forcing of the upper layer, and these anomalies are represented by a stochastic term in the model equation. In the stochastic climate model of chapter 3, this stochastic term represented the synoptic heat flux variability in the surface fluxes. Now, in the 1S model, the focus is on the freshwater flux variability with its typical time scale of a few months. Although heat flux variability plays the dominant role in triggering a single convection event, the freshwater flux variability is an important contribution to the long-term evolution of the background stratification of the water column (Houghton and Visbeck, 2002). Thus, the focus is now on the interannual variability of lateral eddy mixing and advection, which are the main sources for freshwater flux variability. As Hall and Manabe (1997) showed, the stochastic approach holds for sea surface salinity as well, especially on time scales below a few years. (The heat flux variability could be represented in a rudimentary way by adding another stochastic term to y_0. Since this does not change the basic model behaviour [see Wunsch (2002)], we rather opted for keeping the model as simple as possible.)

The 1S model equation (4.25) is thus extended by a noise process ξ with a decorrelation time τ_ξ and a noise intensity σ:

$$\frac{dy}{dt} = -k\,y + k_S(y^* - y) + \xi \tag{4.35}$$

$$\frac{d\xi}{dt} = -\frac{1}{\tau_\xi}\xi + \frac{\sigma}{\tau_\xi}\zeta_t \tag{4.36}$$

$$k = 0 \quad \text{for} \quad y \le y_0$$

$$k = k_c \quad \text{for} \quad y > y_0$$

The noise process is driven by a Gaussian white noise process ζ_t. In the limit $\tau_\xi \to 0$ the noise process $\xi(t)$ becomes white noise as well; with $\tau_\xi > 0$ the process $\xi(t)$ is red noise. Estimations of all parameter values are given in section 4.3.3.

The 1S model and the 2TS model differ in where they separate the "fast" from the "slow" time scales (sect. 4.2.3). In the 1S model, we have assumed an instantaneously relaxed upper box temperature. This means that this relaxation, taking place on a time scale of a few months, is now considered as "fast". By contrast, in the 2TS model the fast time scale in this sense was the synoptic time scale of a few days. With stepping from the 2TS model to the 1S model, the time scale separation border between resolved processes and those parts of the dynamics that are considered as fast shifts from a few days to a few months.

In general, the fast processes are represented by the stochastic forcing. However, in both models the convective mixing is another fast process. The mixing time scale is given by the numerical integration time step in the 2TS model (due to the convective adjustment), and it is of the order of a few days in the 1S model. To have two competing fast processes is not in full agreement with Hasselmann's theory. Yet, from the physics of deep convection it is justified to construct the two models in the way we have done it here: the forcing as well as the convection events are indeed fast processes in the real ocean. Moreover, the results of the previous chapter have shown that in the limit of weak noise, the behaviour of the stochastic model converges to that of the deterministic model.

For the next few sections the noise is assumed to be white ($\tau_\xi = 0$), which makes the analytical computations easier. The effect of red noise, which is more realistic due to its finite decorrelation time, is shown later on.

A typical trajectory of the model (Fig. 4.4) shows how the model fluctuates around the two states, and how it crosses the separating threshold from time to time. The trajectory can be characterized by the stationary probability density function p_s (see section 4.2.1). With the potential (4.32) and (4.33) it is straightforward to use the Fokker-Planck equation in order

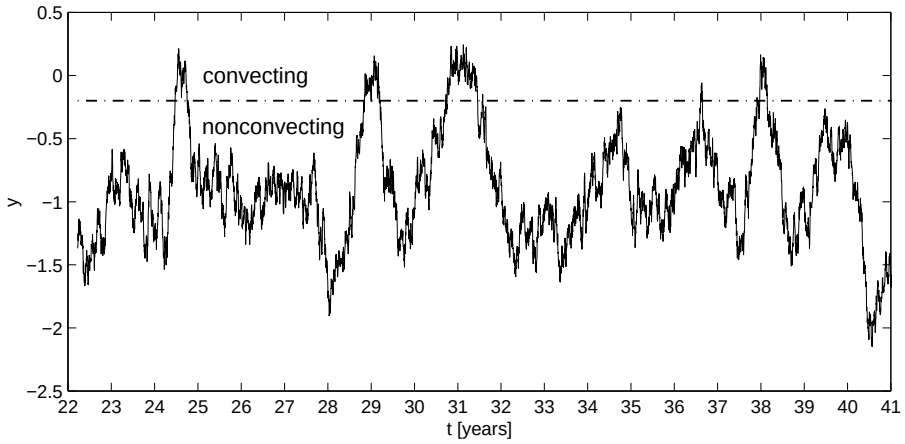

Figure 4.4: *Time series from the 1S model with parameters* $k_S = 1$, $k_c = 10$, $\sigma = 0.8$, $y^* = -1$, *and* $y_0 = -0.2$. *The parameters are the same as for the middle panel of Fig. 4.3; the model is in the bistable domain. The dash-dotted line denotes the threshold* y_0 *that separates the convecting regime* ($y > y_0$) *from the nonconvecting regime* ($y \leq y_0$). *Convection occurs in the model years 24, 28, 29, 30, 31, 36, 37, and 38.*

to determine p_s for (4.35):

$$p_s(y) = \mathcal{N} b_1 \exp\left[-\frac{k_S}{\sigma^2}\,(y - y^*)^2\right] \quad \text{if } y \leq y_0\,, \tag{4.37}$$

$$p_s(y) = \mathcal{N} b_2 \exp\left[-\frac{k_S K}{\sigma^2}\left(y - \frac{y^*}{K}\right)^2\right] \quad \text{if } y > y_0\,. \tag{4.38}$$

There are two abbreviating constants

$$b_1 = \exp\left[\frac{k_S}{\sigma^2}\,(y_0 - y^*)^2\right]\,, \tag{4.39}$$

$$b_2 = \exp\left[\frac{k_S K}{\sigma^2}\left(y_0 - \frac{y^*}{K}\right)^2\right]\,. \tag{4.40}$$

The constant $\mathcal{N}$ is determined by the normalization condition

$$\int_{-\infty}^{\infty} p_s(y)dy = 1\,. \tag{4.41}$$

Three typical probability density functions (pdfs) are shown in Fig. 4.5. From the exponential term in (4.37) and (4.38) it is obvious that the pdf peak is sharper for a deeper well (large k_S values) and for weaker noise

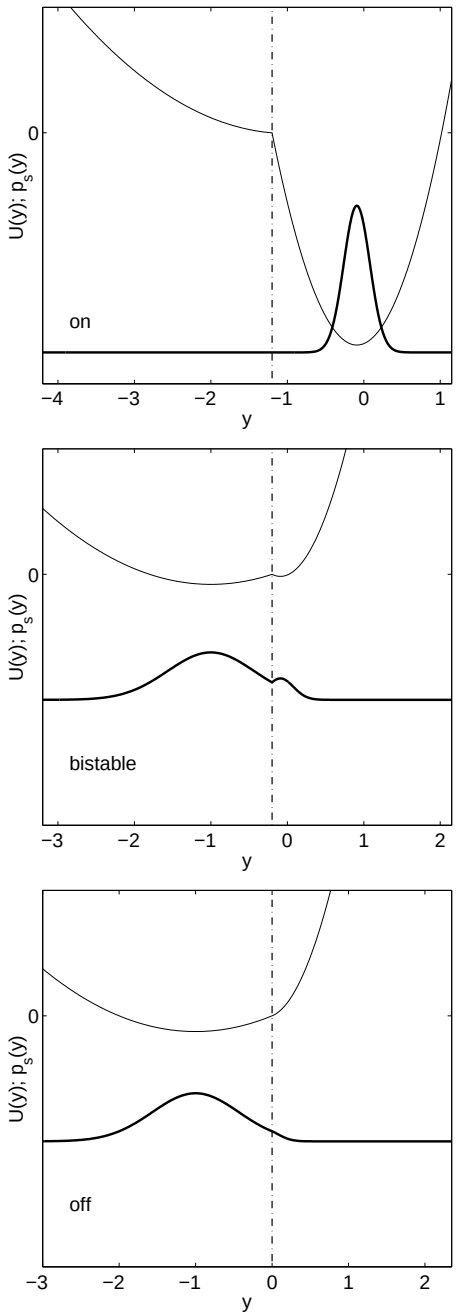

Figure 4.5: *Probability density function p_s (thick, in arbitrary units) corresponding to the three potential curves of Fig. 4.3. The potential curves are repeated here (thin). Parameter values as in Fig. 4.3; the white noise intensity is $\sigma = 0.8$. The dash-dotted line denotes the convection threshold y_0. The two peaks in the bistable case (middle panel) are clearly different in size, indicating that the "off" regime (to the left of the dash-dotted line) is occupied more often than the "on" regime. Conversely, the lower panel shows that there is a non-zero, finite probability for the "on" regime to be occupied even if the stable "on" state does not exist.*

(small σ values). In the deterministically monostable domains the pdf has one peak, and there are excursions over the threshold into the other regime. Stronger variability leads to more excursions over the threshold. If the deterministic system is bistable, then the pdf has two peaks. Yet, the different size of the two peaks indicates that the trajectory spends more time in one regime than in the other. If the peak sizes are very different, this may mean that one regime is reached only rarely.

Neglecting transients due to particular initial conditions, the probability for the model to be in the convecting regime is given by

$$P_c = \int_{y_0}^{\infty} p_s(y)dy\,. \tag{4.42}$$

Similarly, the probability for the model to be in the nonconvecting regime is

$$P_n = \int_{-\infty}^{y_0} p_s(y)dy\,. \tag{4.43}$$

These two probabilities are useful in later sections.

The results from the previous chapters motivate to study the model behaviour in the "bistable" and in the "off" domain. In contrast, Cessi (1996a), although using a model very similar to ours, studied the behaviour in an oscillatory state of the model. The observed temperature and salinity stratifications in the North Atlantic never lead to such an oscillatory state; Cessi (1996a) aimed at analyzing oscillations of the THC occurring in GCMs.

4.3.3 Parameter estimation

The 1S model draws a highly idealized picture of convection dynamics. It is designed as a tool to study the dynamics of the 2TS model, but it is not meant to realistically model salinity time series by itself. A detailed parameter estimation like the one carried out for the 2TS model is thus not feasible. We will work with two parameter sets: one "tutorial" that helps to clarify the basic model properties, and one "estimated" from observational data that is somewhat more realistic.

Considering the "estimated" parameters first, we start with taking over the parameters from the 2TS model (table 2.2) and derive the 1S parameters according to eqs. (4.27) to (4.29). This yields $y^* \approx -1.5$ psu and $y_0 = 0.04$ psu. With the time unit being still one year, we have $k_S = 0.125\,\mathrm{yr}^{-1}$ from $\tau_{1S} = 8$ yr. The convective mixing is assumed to have $k_c = 50\,\mathrm{yr}^{-1}$, corresponding to a time scale of one week. Larger values of k_c would require numerical time steps smaller than 1 day, which is undesirable with respect to the time scale of the stochastic forcing. For the sake of brevity, the units of the parameters are left out hereafter. The "estimated" parameter set is thus $k_S = 0.125, k_c = 50, y^* = -1.5$, and $y_0 = 0.04$.

A check for this parameters is possible with data from Houghton and Visbeck (2002), hereafter cited as HV02. They specify a value of 0.2 Sv total mean freshwater flux into the Labrador Sea (1 Sv = 10^6 m^3s^{-1}). Most of this freshwater is transported by the boundary currents, such that only 20%, or 0.04 Sv, reach the interior Labrador Sea through lateral eddy mixing. They further assume a volume of $V = (0.6 \cdot 10^6\,\mathrm{km}^2) \cdot 300\,\mathrm{m}$ for the interior Labrador Sea, and a reference salinity of $S_0 = 35$ psu. One obtains a mean freshwater flux of

$$\Phi_{FW} = \frac{S_0}{V} \cdot 0.04\,\mathrm{Sv} = 0.25\,\mathrm{psu\,yr}^{-1}\,. \tag{4.44}$$

According to HV02 the error of this estimated freshwater flux is 50%. Thus, the value of Φ_{FW} is consistent with the observation that, if convection is absent, the upper layer salinity decreases with about half of this rate (table 2.1, see also Fig. 5 of HV02). In the 1S model, the initial salinity decrease immediately after the end of convection ($y(t_0) = y_0$) is

$$\frac{dy}{dt} = k_S(y^* - y(t_0)) = 0.125 \cdot (-1.5 - 0.04)\,\mathrm{psu\,yr}^{-1} \approx -0.2\,\mathrm{psu\,yr}^{-1}\,, \tag{4.45}$$

which is consistent with the above value of Φ_{FW}, too.

However, there is a difference to the "flushing time" of the interior Labrador Sea as estimated by Khatiwala and Visbeck (2000). From dynamical considerations, they obtain a value of 3 yr for the interior Labrador Sea to be ventilated by lateral mixing and deep convection. Our value of $\tau_{1S} = 8$ yr (or $k_S = 0.125$) is somewhat larger. Yet, their study has the baroclinic lateral mixing as a prerequisite. The presence of deep convection is needed to maintain the horizontal salinity gradient that drives the lateral eddy mixing. If convection is absent for a longer time, then freshwater accumulates in the interior Labrador Sea. Thus this horizontal gradient is removed, and the lateral mixing becomes weaker (Dickson et al., 1988). This fits with the box model's relaxation being strong in the beginning of a non-convecting phase, but lessening later on. Whereas the flushing time of Khatiwala and Visbeck (2000) was obtained for a phase with frequent convection, our box model applies also to long nonconvecting phases.

Due to the absence of the seasonal cycle the estimated parameters reflect the annual mean state. With $y_0 > 0$, the model is located in the "off" domain. Only a forcing with a magnitude similar to the amplitude of the seasonal cycle will then be able to trigger convection events. If we do not want to include the seasonal cycle itself (to keep the model simple), then the stochastic forcing term has to take over this amplitude. Alternatively one could derive the parameters from permanent winter conditions. The stochastic forcing would then represent the interannual variability only. The results are affected only quantitatively by this alternative.

We still need to estimate the decorrelation time τ_ξ and the standard deviation std(ξ) of the noise term ξ in (4.35). The decorrelation time scale

of the freshwater flux is difficult to estimate from observations because long time series are sparse, and the freshwater flux is fed from many sources (such as continental runoff, sea ice advection and melt, precipitation). Proxy time series from models suggest to assume a decorrelation time of half a year to three years. Such time series are, for instance, the sea ice export through Fram strait (H. Haak, pers. comm.) or the sea ice volume in Baffin Bay (M. Maqueda, pers. comm.). We use here $\tau_\xi = 2$ years. In this way, the stochastic freshwater forcing includes interannual anomalies.

According to HV02, the anomalous freshwater flux associated to a Great Salinity Anomaly is 20% of the mean freshwater flux. This yields a standard deviation $\mathrm{std}(\xi) = 0.05\,\mathrm{psu\,yr}^{-1}$ for the interannual variability. The anomalous freshwater flux amounts to 20% of the processes that drive the seasonal cycle (HV02). If the stochastic forcing term is to contain the variability of the seasonal cycle as well, then a sensible choice is $\mathrm{std}(\xi) = 0.25\,\mathrm{psu\,yr}^{-1}$. Since we will also use white noise forcing, it is useful to determine the noise intensity σ in (4.36) from $\sigma = \sqrt{2\tau_\xi}\,\mathrm{std}(\xi)$ (Gardiner, 2002). With the above values this yields $\sigma = 0.5\,\mathrm{psu\,yr}^{-1/2}$.

Parameter set	k_S (yr^{-1})	k_c (yr^{-1})	σ ($\mathrm{psu\,yr}^{-1/2}$)	τ_ξ (yr)	y^* (psu)	y_0 (psu)
"estimated"	0.125	50	0.5	2.0	-1.5	0.04
"tutorial"	1.0	10	0.8	0	–	–

Table 4.1: *Parameter sets for 1S model simulations. For the tutorial parameters no particular values for y^* and y_0 are specified.*

The "estimated" values for k_S and k_c result in two strongly asymmetric potential wells. The "on" well is very narrow and would be hard to visualize. For tutorial reasons, we have therefore chosen $k_S = 1$, $k_c = 10$, and $\sigma = 0.8$ for the next few sections, and return to the estimated parameter values later. These "tutorial" parameters still capture the two different regimes and are not completely unphysical. Moreover, with the tutorial parameters we use a stochastic forcing with $\tau_\xi = 0$ (white noise), to make the analytical computations easier. Both parameter sets are given in table 4.1. Additionally, the values $k_S = 10$ and $k_c = 50$ were used in two cases for illustrative reasons.

4.4 Residence times

4.4.1 Mean residence times

The mean residence times are a relevant dynamical feature of a stochastic dynamical system. They give the information how often the jumps between

the regimes occur, while the probabilities P_c and P_n only tell how much time the system has spent in total in either regime. The mean residence times are also important for defining stochastic stability (section 4.2). For the 1S model an analytical approximation for the mean residence time is derived. A comparison with numerical computations will show the limits of this approximation.

The residence time t_r is defined here as the time that a trajectory spends uninterruptedly in one regime. Inspection of Fig. 4.4 shows that this definition includes long residences of several years where the trajectory spends much time close to the deterministic stable state – but also short residences of a few time steps' length where the trajectory stays close to the threshold y_0. Defining the residence time in this way is appropriate in the context of the 1S model, since we are interested in the transitions between the convecting an the nonconvecting regimes. In particular, it does not matter whether one of the stable states does not exist. The definition relies only on the regime transitions.

Another approach is needed for an approximate analytical calculation of the residence times. Here an expression for the mean *escape* time can be obtained. Provided that a deterministically stable state exists, the escape time t_e is defined as the time the trajectory spends in the corresponding pdf peak before it hits a given threshold the first time. This definition is commonly used in physics (Hänggi et al., 1990). For a system moving in a potential U, we know (eq. 4.15) that the mean escape time $\langle t_e \rangle$ from a potential minimum (here at y^*) to a threshold (here at y_0, and $y_0 > y^*$) is

$$\langle t_e \rangle = \frac{2}{\sigma^2} \int_{y^*}^{y_0} \exp\left[\frac{2}{\sigma^2} U(y)\right] \left(\int_{-\infty}^{y} \exp\left[-\frac{2}{\sigma^2} U(y')\right] dy' \right) dy \,. \tag{4.46}$$

This equation is now used to obtain the mean escape time $\langle t_{e,n} \rangle$ from the nonconvecting state. If we assume small noise, then the first exponential is large only close to y_0, while the second has significant magnitude only around y^*. Hence the contribution from the second integral is relevant for y close to y_0 only, and will not vary strongly for these y values. Therefore we can set y to y_0 in the upper bound of the second integral and treat the two integrals as approximatively independent. Further, to solve the first integral we linearize $U(y)$ around y_0. With these assumptions, the solution of (4.46) is

$$\langle t_{e,n} \rangle \approx \frac{\sqrt{\pi}}{k_S} \left(\frac{k_S}{\sigma^2} (y_0 - y^*)^2 \right)^{-\frac{1}{2}} \exp\left[\frac{k_S}{\sigma^2} (y_0 - y^*)^2 \right] . \tag{4.47}$$

The mean escape time from the nonconvecting state predominantly depends on the exponential of the potential well depth divided by the squared noise intensity. Equation (4.47) is a generalization of a result already obtained by Kramers (1940) for the escape time from a parabolic potential well over a kink in the potential curve into a second, symmetric potential well (see also Hänggi et al. (1990)).

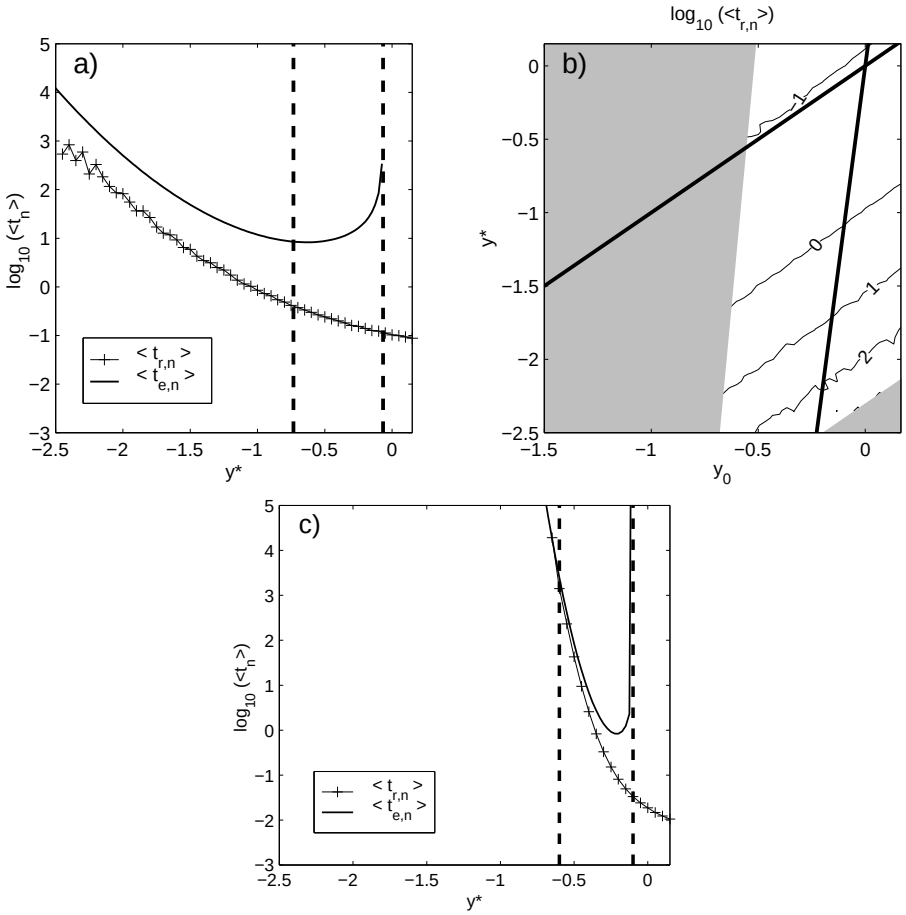

Figure 4.6: *(a) Comparison of the mean residence time in the nonconvecting regime* $\langle t_{r,n} \rangle$ *(logarithmically scaled, time unit is one year) from simulations (crosses) with the analytically computed mean escape time* $\langle t_{e,n} \rangle$ *(solid) as a function of* y^* *for fixed* $y_0 = -0.067$*. The dashed lines enclose the bistable domain. (b) Contour plot of the mean residence time (logarithmically scaled) as a function of the two model parameters* (y_0, y^*)*. Due to the finite length of the model simulations the contours are not perfectly smooth and could not be computed for the whole parameter plane. Shading indicates where they miss. The thick lines are the deterministic stability borders like in Fig. 4.2. For panels (a) and (b) the tutorial parameters were used. (c) shows the same comparison like in (a), but with other parameters:* $k_S = 10, k_c = 50, \sigma = 0.5$*, and* $y_0 = -0.1$*. Note that the analytical approximation in (c) holds much better because of the larger value of* k_S*. The deterministic stability borders (thick line and/or dashed lines) and the shading will appear in some of the following plots.*

In comparing our result (4.47) with an estimate of the mean residence time from numerical simulations (Fig. 4.6), several points deserve our attention.

- The analytical expression (4.47) for the mean escape time $\langle t_{e,n} \rangle$ is a good approximation to the mean residence time $\langle t_{r,n} \rangle$ if there is a sharp pdf peak at the "off" state. This is only the case for $y^* \leq y_0$, and moreover requires the term $k_S \, (y_0 - y^*)^2 \, \sigma^{-2}$ to be large (cf. eq. 4.37). This is achieved by either a strong restoring coefficient k_S, small noise σ or a large difference $(y_0 - y^*)$, which is seen in comparing Fig. 4.6a with Fig. 4.6c. For a small difference $(y_0 - y^*)$, the term $(k_S \, (y_0 - y^*)^2 \, \sigma^{-2})^{-1/2}$ in (4.47) grows large and the approximation breaks down (right-hand part of Fig. 4.6a).

- The large difference between the analytical $\langle t_{e,n} \rangle$ and the numerical $\langle t_{r,n} \rangle$ in Fig. 4.6a as well as the large excursions of the time series in Fig. 4.4 demonstrate that the noise intensities used here are rather high. At the same time, we see from the left-hand part of Fig. 4.6a that the analytical results are an important guidance to understand the model behaviour, even if the noise is not "small" in the sense of the assumption that underlies the analytical calculations.

- The analytical approximation for the mean *escape* time $\langle t_{e,n} \rangle$ is valid only for those events where the system moves from the pdf peak to the threshold. The numerical mean *residence* time $\langle t_{r,n} \rangle$ is more comprehensive: it includes as well short back-and-forth crossings of the threshold. These events lead to smaller values for $\langle t_{r,n} \rangle$ than for $\langle t_{e,n} \rangle$. Those short crossings however produce a time-step dependence of $\langle t_{r,n} \rangle$: the smaller the time step, the smaller is the shortest possible residence time. This undesirable feature is overcome by introducing the coarse-grained statistics in section 4.5.

- The mean residence time $\langle t_{r,n} \rangle$ is a smooth function of the model parameters y_0 and y^* irrespective of the deterministic stability domain borders. The right-hand dashed line in Fig. 4.6a denotes the y^* value where the nonconvecting state ceases to exist, but the curve of the mean residence time does not show a sign of this particular point. The mean residence times do not show abrupt changes in dependence on the model parameters, in contrast to the deterministic stability properties. Although not shown here, this is also true for the distribution of residence times.

- The analytical approximation shows that the mean residence times depend exponentially on the forcing parameters like the vertical temperature gradient y_0 or the reference vertical salinity gradient y^*. The

numerical simulations show that this holds true throughout the parameter space, independent of the deterministic stability of the model states, and even if the analytical approximation breaks down.

- Since the mean residence time in one regime depends on the potential function in this regime, the lines of equal $\langle t_{r,n} \rangle$ are parallel to isolines of the potential function in the nonconvecting regime (Fig. 4.6b).

These properties of the mean residence time are crucial for the stochastic stability concept developed later on.

We note that Cessi (1994) computed the mean residence times in a similarly simple model. She simplified Stommel's (1961) box model of the thermohaline circulation to a version with one single variable, the horizontal salinity gradient in the Atlantic. The residence times were computed for two states, a weak and a strong meridional overturning circulation. While her focus was to characterize the variability of the large-scale circulation, we aim at analyzing the local stability of the (potentially) convecting water column.

4.4.2 Probability of residence times exceeding a threshold

Having considered the mean residence time in a regime, we now turn towards the distribution of the residence times and the probability for the residence time to exceed a threshold value. In general, for a random process with vanishing autocorrelation it is known (Leadbetter et al., 1983; von Storch and Zwiers, 1999) that the residence times t_r in a given regime are exponentially distributed. The probability density function is then:

$$p_s(t_r) = \frac{1}{\langle t_r \rangle} \exp\left[-\frac{t_r}{\langle t_r \rangle}\right], \tag{4.48}$$

with $\langle t_r \rangle$ the mean residence time. The probability distribution for t_r reads:

$$P(t_r < t_x) = \int_0^{t_x} p_s(t_r) dt_r = 1 - \exp\left[-\frac{t_r}{\langle t_r \rangle}\right] \tag{4.49}$$

The function $P(t_r < t_x)$ gives the probability for the residence time to stay below a threshold t_x. It is easy now to compute, for instance, the probability of the residence time $t_{r,n}$ to be larger than one year. The only required quantity is the mean value $\langle t_{r,n} \rangle$:

$$P(t_{r,n} > t_{yr}) = 1 - P(t_{r,n} < t_{yr}) = \exp\left[-\frac{t_{yr}}{\langle t_{r,n} \rangle}\right] \tag{4.50}$$

Yet, the noise process considered here, the output from the 1S model, has a considerable autocorrelation that is determined by k_S and k_c. In this case

the distribution of $t_{r,n}$ deviates from the ideal exponential form, as seen in section 4.6. An analytical treatment is still possible (Leadbetter et al., 1983), but tedious. Therefore (4.50) is used for further calculations, but only as an approximation.

4.5 Coarse-grained statistics of the model time series

For the sake of simplicity we have avoided to explicitly introduce the seasonal cycle. Instead we let the stochastic forcing take over its amplitude. Deep convection events now may occur at any time in a given year, and not only during the cycle's extremum in winter. This is a problem for counting convection events. Introducing a coarse-grained criterion to interpret the time series from the 1S model provides a way out.

In the real ocean the physically most relevant feature of a deep convection event is the mixing of the deep waters with the surface layer waters. As the diffusive and advective time scales of the deep ocean are clearly larger than one year, and as there is virtually no seasonal cycle in the deep waters, it does not really matter when exactly in a given year deep convection occurs. Under constant boundary conditions, the relevant physical question is whether convection did occur in a given year at all. Therefore, a given year of the model output is called a "convecting year" if there was at least one convection event during this year. This rule establishes the coarse-grained criterion. As an illustrative example, consider the time series displayed in Fig. 4.4. The model years 24, 28, 29, 30, 31, 36, 37, and 38 are "convecting years" in the above sense.

The question how often convection occurs can be answered by the probability n_c for a convecting year to occur in a model run. The probability n_c is coarse-grained in the sense that it neglects the short back-and-forth changes between the convecting and the nonconvecting regime. Although these short changes may occur in the real ocean, it is the cumulated effect of all those short convection events that determines water mass properties and the circulation in the ocean. In contrast to the mean residence time, n_c depends only very weakly on the time step size. Changing the time step by one order of magnitude leads to changes in n_c that hardly exceed 10%. In any case, it is justified to use the coarse-grained probability n_c as long as the focus is on the statistics of deep convection events.

An analytical expression for n_c is readily derived. Suppose that the nonconvecting stable state exists so that (4.47) is valid. The time span of one year is called t_{yr}. (The definition of "year" is arbitrary here, it could be any time interval of fixed length.) Obviously,

$$n_c = 1 - P(y < y_0 \text{ during } t_{yr})\,. \tag{4.51}$$

The probability $P(y < y_0 \text{ during } t_{yr})$ for convection not to occur in a year is the probability to be in the nonconvecting regime in the beginning of this year times the probability for the residence time in the nonconvecting regime to exceed one year:

$$P(y < y_0 \text{ during } t_{yr}) \tag{4.52}$$

$$= P(y(t_0) < y_0) \cdot P(y(t) < y_0 \text{ for all } t_0 < t \leq t_0 + t_{yr}) \tag{4.53}$$

$$= \int_{-\infty}^{y_0} p_s(y)dy \cdot \exp\left[-\frac{t_{yr}}{\langle t_{r,n}\rangle}\right] . \tag{4.54}$$

Finally,

$$n_c = 1 - P_n \cdot \exp\left[-\frac{t_{yr}}{\langle t_{r,n}\rangle}\right] . \tag{4.55}$$

The probability n_c is a function of the probability to be in the nonconvecting regime and of the mean residence time in this regime.

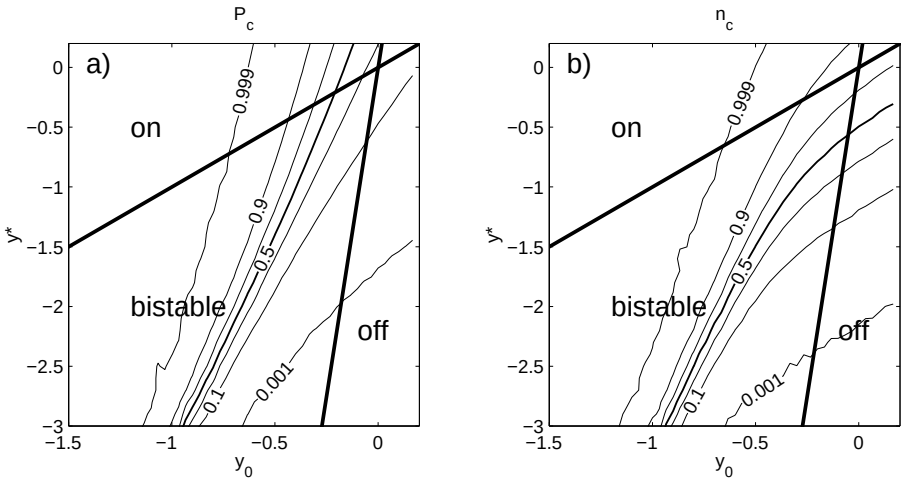

Figure 4.7: *Estimates of (a) the probability P_c to be in the convecting regime and (b) the probability n_c for a convecting year from numerical simulations of the 1S model. Contours show P_c and n_c as a function of the forcing parameters y^* and y_0, using the tutorial parameters. Thick lines denote the deterministic stability domain borders as in previous figures. Note that the contours run smoothly through these deterministic stability borders.*

It is instructive to compare n_c, the probability for a convecting year, with P_c, the probability to be in the convecting regime. Figure 4.7 shows the differences. Consider first the lower left-hand part of the panels. This is the bistable domain with $y^* < -1.5$, say. The values of P_c and n_c are very similar. Both potential wells are rather deep here, and the mean residence

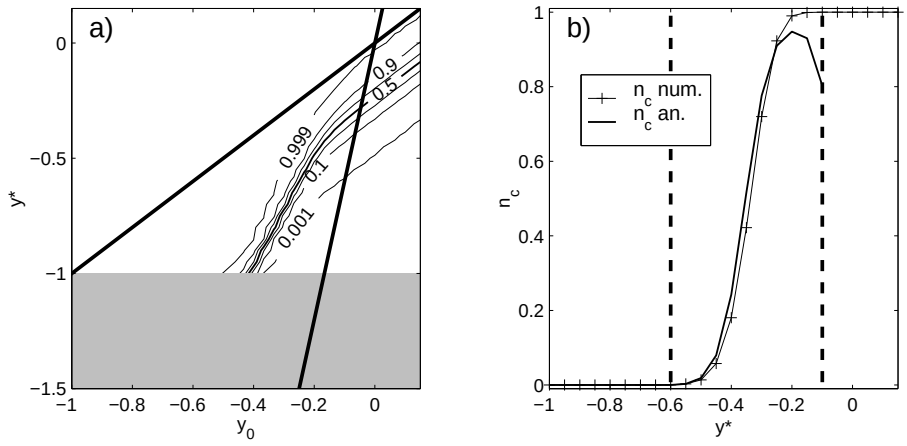

Figure 4.8: *(a) Contours of n_c for $k_S = 10, k_c = 50$ and $\sigma = 0.5$. For small y^* values the contours follow the $\langle t_{r,n} \rangle$ contours, or the isolines of the potential function of the "off" regime. (b) Comparison of numerical (crosses) and analytical (solid line) computation of n_c for the same parameter values with fixed $y_0 = -0.1$. The analytical approximation breaks down where the analytical $\langle t_{e,n} \rangle$ has its minimum (cf. Fig. 4.6c).*

times are large. Consequently, the exponential in (4.55) is close to one, so in the limit we have $n_c = 1 - P_n \cdot 1 = P_c$. Now consider the upper right-hand part of the panels. This is the "off" domain with $y^* > -1.5$ or so. Here the differences of P_c and n_c are considerable. The P_c values are small since the stable convecting state does not exist. In contrast, the n_c values are large and even exceed 0.5. The convecting state does not exist, but the probability for a convecting year can be larger than 50%. The explanation lies in the "off" potential well being shallow here. The small residence time means that there are frequent excursions from the non-convecting equilibrium state where the model hits the convection threshold. The pdf peak is broad, as seen in the "off" panel of Fig. 4.5. Since $P_n \approx 1$, we have, again in the limit, $n_c = 1 - 1 \cdot \exp[-t_{yr}/\langle t_{r,n} \rangle]$. In the "off" domain, n_c depends on the mean residence time only. This feature comes out more clearly in Fig. 4.8a. For large y^* the isolines of n_c are parallel to the "off" domain borders. These parallels are the lines of equal $(y_0 - y^*)$ and therefore lines of equal shape of the potential function in the nonconvecting regime.

With (4.55) n_c can be computed analytically. The mean residence time obtained numerically is then replaced by the analytical expression for the mean escape time (4.47). In addition, the n_c definition relies on the assumption of exponentially distributed residence times. Fig. 4.8b shows a comparison of the analytically and the numerically computed n_c. The analytical approximation obviously follows the numerical values quite closely

over a large range of y^* values. This is an example of how successful the 1S model is in providing conceptual and analytical approaches to understand the statistics of deep convection events.

The qualitative differences between P_c and n_c are the same when not using the "tutorial" parameters, but the more realistic "estimated" parameters. Fig. 4.9 shows that with the estimated parameters, the dependence of P_c and n_c on y^* is weaker than in Fig. 4.7, but the dependence on y_0 is stronger. The reason lies in the ratio of the two potential wells in either parameter set. A convenient measure for this ratio is K (see (4.32) and (4.33)): for the tutorial parameters $K = 11$, whereas the estimated parameters have $K = 401$. In any case, in the "off" domain the values of n_c are still larger than the P_c values.

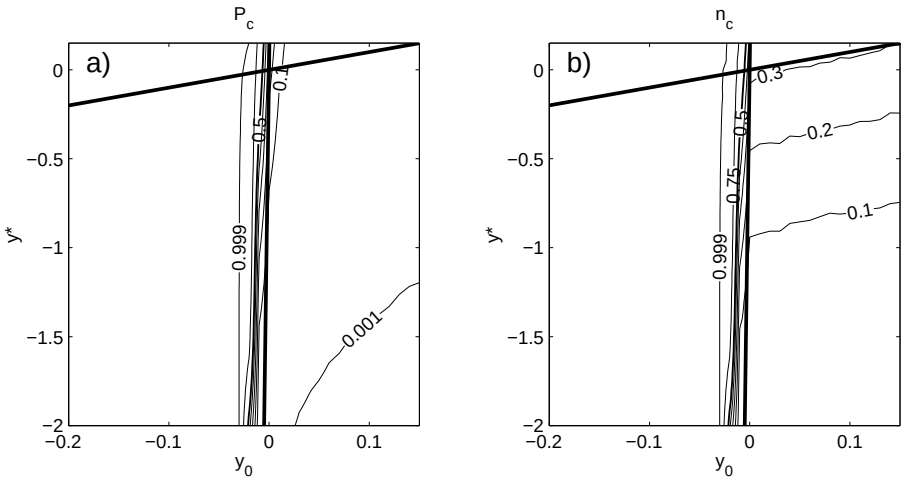

Figure 4.9: *Contours of (a) P_c and (b) n_c for the estimated parameter set. The difference between P_c and n_c is still significant, particularly in the "off" domain ($y_0 > 0$).*

The differences between P_c and n_c are remarkable in their physical interpretation. The frequent occurrence of convecting years does not depend on the existence of the convecting state, because short excursions over the convection threshold into the convecting regime are sufficient to achieve convection events. The curves of n_c run smoothly across the point where the convecting stable state ceases to exist.

4.6 Nonconvecting phases

We consider now the probability for uninterrupted sequences of convecting years (or of nonconvecting years). Each of these sequences is called a (non)convecting *phase*. We compare the results of the 1S model with those

of the 2TS model. Fig. 4.10 shows the distributions of the length of the two kinds of phases. The distribution of the frequency of the phase lengths may again be interpreted as a pdf of the phase length. Obviously, Fig. 4.10 is qualitatively similar to the Figure from the 2TS model (Fig. 3.4). Again, the convecting phase lengths are distributed almost exponentially, indicated by the almost straight line of the distribution in the logarithmically scaled plot (panel a). In contrast, the distribution of the nonconvecting phase lengths is bent, with short phases occurring rather often, but also a small frequency of nonconvecting phases that are longer than 100 yr (not shown here).

At this point the positive salinity feedback comes into play again. Like in the 2TS model, this feedback acts during the nonconvecting phases, rendering their termination less and less probable the longer they last. This

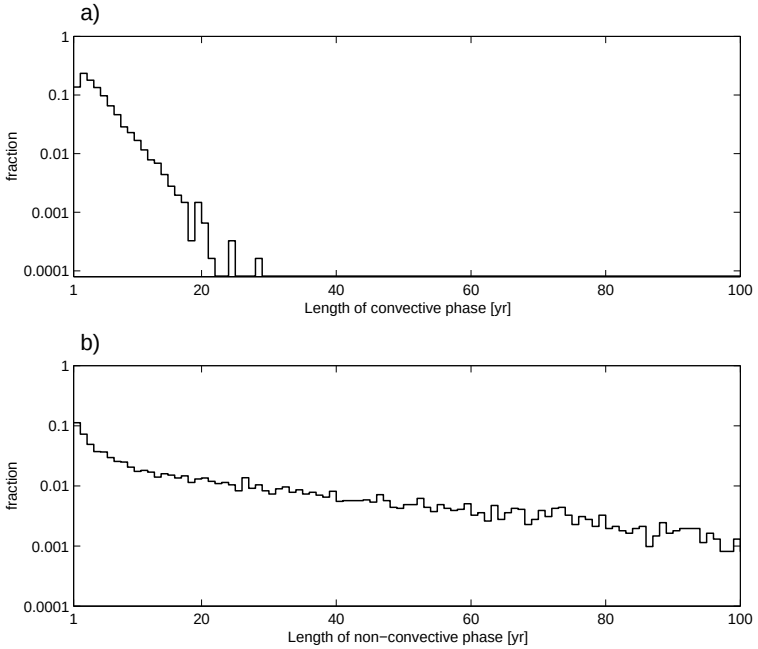

Figure 4.10: *Distributions of the lengths of convecting phases (a) and nonconvecting phases (b), where a convecting phase is an uninterrupted sequence of convecting years. The frequency of every single phase length is given as a fraction of the total number of phases during a* 10^5 *yr model run. The fraction axis is logarithmically scaled. For the analyzed model run the estimated parameters were used (but* $y^* = -0.75$*). The model is in the (deterministically) monostable "off" domain. The distribution in (b) shows the characteristic "red-noise bent" at short phase lengths. The mean phase lengths are 4.0 yr (a) and 29 yr (b). The tail of the distribution in (b) was cut off arbitrarily.*

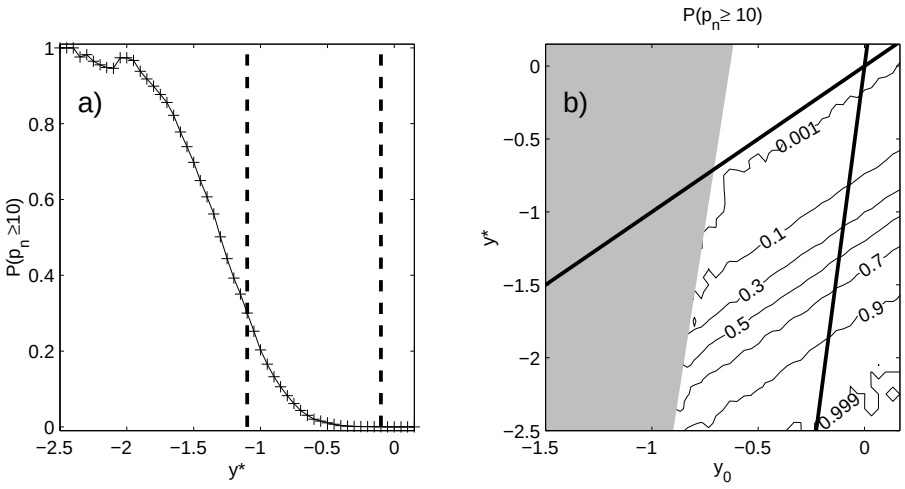

Figure 4.11: *Probability for the length p_n of the nonconvecting phase to be 10 yr or longer: (a) as a function of y^* with fixed $y_0 = -0.1$, (b) contour plot as a function of y_0 and y^*. Tutorial parameter set. Shading, thick and dashed lines as in Fig. 4.6.*

is the reason for both the relatively high probability for short phases and the small, but non-negligible probability for very long phases. The point is that this feedback is acting in the model run analyzed in Fig. 4.10, although parameters from the monostable "off" domain were used. In other words, the positive salinity feedback is not tied to the bistable domain. Instead, it is the "memory" of the system being longer than one year (namely, $1/k_S = 8$ yr) that produces the feedback. This is easily seen when looking at the phase length pdf from model runs with $k_S = 1$ (not shown here). In this case, the dynamical time scale is $1/k_S = 1$ yr. There is no long memory in the system, and the nonconvecting phase lengths are distributed exponentially as well. The positive salinity feedback is represented by the construction of a sufficiently broad single potential well plus the convection threshold; the bistability is not required.

The pdf can be cumulated to give the probability for the phase length p_n to exceed a certain threshold (cf. Fig. 3.5). Here we are interested in how this probability depends on the forcing parameters. As an example, Fig. 4.11a shows how the probability $P(p_n \geq 10)$ for a nonconvecting phase to last 10 yr or longer depends on the value of the reference salinity gradient y^*. Where this probability is very small for y^* close to zero, corresponding to saline surface conditions, it increases with a fresher surface forcing. For $y^* = -1$ its value is about 0.2, meaning that, when a nonconvecting phase starts, chances are 20% that it will last 10 yr or even longer. If y^* is further decreased, $P(p_n \geq 10)$ continues to grow and eventually reaches 1. The

whole curve has a shape similar to a hyperbolic tangent. Equation (4.50) reveals that $P(p_n \geq 10)$ depends on the mean residence time only. Thus its contours are parallel to the line $y_0 = y^*$ (Fig. 4.11b), as for n_c in the "off" domain (Fig. 4.8a).

Note that $P(p_n \geq 10)$ shows an exponential dependence on the parameters if its values are small. However, for $0.2 < P(p_n \geq 10) < 0.8$ the dependence is almost linear. If we interpret the fact that a nonconvecting phase lasts 10 yr or longer as a kind of a climatic extreme event (which it would be at least in the Labrador Sea), then this result is an important extension to the findings of Khatiwala et al. (2001).

Using a model very similar to our stochastic 1S model, Khatiwala et al. (2001) analyzed how the slope of the phase length pdf (i.e. the slope of a straight line fit to Fig. 4.10a) depends on a forcing parameter (their Fig. 4a). It turned out that this is a linear dependence. Thus, the parameter governing the exponential distribution of the phase lengths is a linear function of the model parameters. Interpreting their simple model as paradigmatic for climatic regimes in the atmosphere, they stated that changing climate parameters might lead to "exponential changes in the occurrence of persistent events" in the atmosphere.

Our analysis however makes it clear that the probability for "persistent events", namely long phases showing no convection, may well be a linear function of the climate parameters if the probability itself is not too small. This is a contrast to the *mean* phase length which always depends exponentially on the parameters. The simplicity of the 1S model – a potential well, a relevant threshold, and stochastic forcing – suggests that our results may be valid as generally as the double-well model of Khatiwala et al. (2001). Whether the exponential rise of the mean phase length or the linear rise of the probability for long phases is the more appropriate perspective depends on the system under consideration. If we study, for instance, the probability for atmospheric regimes yielding long droughts for certain regions, then already the linear increase in probability might be critical and threatening. In any case, our view that includes the *distribution* of phase lengths is more comprehensive than only considering their mean.

4.7 Stochastic stability in comparison with deterministic stability

4.7.1 Quantitative stochastic stability and effective monostability

The framework of the simple 1S model allows to study the stochastic stability measures developed in section 4.2.2 in detail. Applied, for instance, to the nonconvecting state of the 1S model, the three stability measures are

1. the probability P_n to be in the nonconvecting regime (eq. 4.43);
2. the mean residence time $\langle t_{r,n} \rangle$ in the nonconvecting regime;
3. the potential difference $\Delta U^{(n)} = |U(y_{st}^{(n)}) - U(y_0)| = |U(y_{st}^{(n)})|$ needed to escape from the nonconvecting state. Often, $\Delta U^{(n)}$ is called the potential well depth.

These measures may be defined analogously for the convecting state. The definitions rely on the existence of the deterministic potential well. In addition, it is assumed that the mean residence time is a time scale of interest, such that transitions from one regime to the other are observed. On time scales smaller than the mean residence time, such transitions are very unlikely. The system stays in one of the two regimes without "feeling" the bistability.

The stochastic stability measures now open the possibility to quantitatively compare the stochastic stability of the two states in the bistable domain. Section 4.2.2 has shown that the potential well depth, if available, is the most convenient stability measure because it does not depend on the noise intensity. To compare the stability of the two states, one can use the potential well depth difference

$$\delta\Delta U := \Delta U^{(n)} - \Delta U^{(c)} = \left| \frac{k_S}{2} \left(-(y_0 - y^*)^2 + K \left(y_0 - \frac{y^*}{K} \right)^2 \right) \right| . \quad (4.56)$$

With the relation (4.14) it is easy to see that the potential well depth difference $\delta\Delta U$ is equivalent to the difference δt_r of the logarithms of the mean residence times

$$\delta t_r := \ln\langle t_{r,n} \rangle - \ln\langle t_{r,c} \rangle = \frac{1}{\sigma^2} \, \delta\Delta U \, , \quad (4.57)$$

where we assumed that the mean residence times do not deviate too much from the mean escape times used in (4.14). The third possibility to compare the stochastic stability of the two states is to use $P_c = 1 - P_n$. The probability to be in one regime is just the fraction of the total time that the system spends in this regime.

The three measures for comparing the stochastic stability are shown in Fig. 4.12c, e, and a. Not surprisingly, they give similar results. For a symmetric potential, the lines of $\delta\Delta U = 0$, $\delta t_r = 0$, and $P_c = 0.5$ would coincide. The asymmetry of the 1S model explains the deviations. The quantities P_c and δt_r can be used as stability measures only in the bistable domain. However, their contours run smoothly across the deterministic stability borders. In this sense, a stochastic system has no notion of the deterministic stability loss. If one considers the system behaviour under

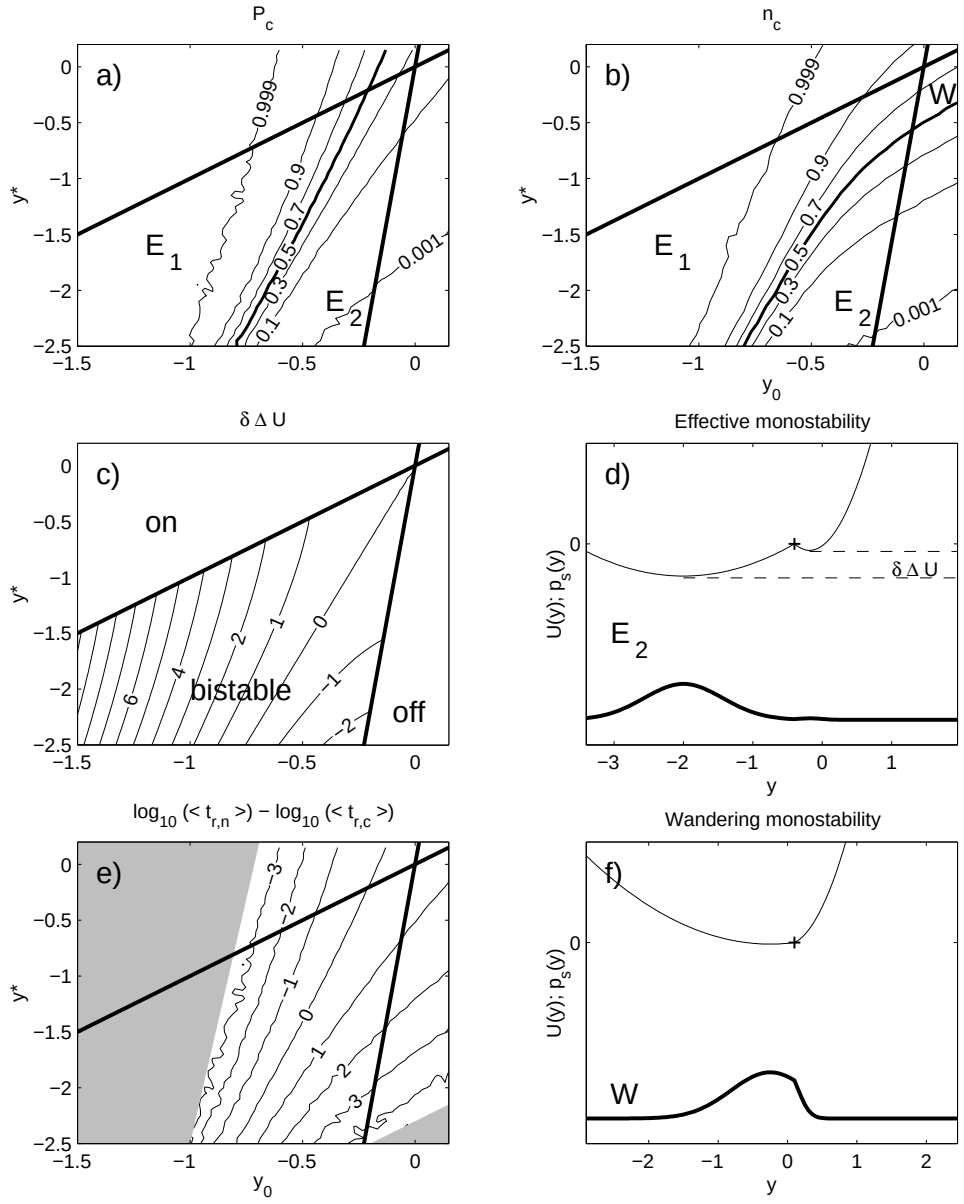

Figure 4.12: *Stochastic stability measures, effective monostability, and wandering monostability. Panel (a) and (b) show P_c and n_c like in Fig. 4.7. P_c is compared with two other stability measures: the potential well depth difference (c) and the difference δt_r of the logarithm of the mean residence times (e). The tutorial parameters are used for all panels. Shading in (e) as in previous figures. Two pdf peaks with strongly different size lead to effective monostability (d), whereas a single pdf peak that leaks into the other regime is associated with wandering monostability (f). Panels (d) and (f) show the potential (thin) and the pdf (thick) as a function of y, with y^* and y_0 corresponding to the position of the letters E_2 and W in (b).*

parameter changes, then a possible sudden loss of stability in the deterministic system is replaced by a continuous tendency to one regime being occupied less often. Consequences of this effect were studied in section 3.4.

Note that there is only a narrow streak in the (y_0, y^*) parameter plane where the probabilities to be in either regime do not differ by at least one order of magnitude. This is consistent with the narrow streak where $|\delta t_r| < 1$. In the largest part of the bistable domain, to either side of that streak, the probability to be in one regime is very close to one, and the other regime is rarely visited. Although the system has two potential wells, one well is much deeper such that the residence time becomes very large. The deterministic bistability is turned into *effective monostability*. The points E_1 and E_2 in the panels of Fig. 4.12 denote the domains of effective monostability, and panel d illustrates the grossly differing size of the two pdf peaks. Effective monostability is a known feature of stochastic bistable systems (Gardiner, 2002), and it has recently been studied in a box model of the thermohaline circulation (Monahan, 2002).

4.7.2 Wandering monostability due to the convection threshold

If one considers the coarse-grained probability n_c, the probability for a convecting year, then another stochastic stability effect comes up. It has been shown that in parts of the "off" domain, the probability for a convecting year is still high, although the convecting state does not exist. The high values of n_c are explained by short excursions of the model trajectory into the convecting regime. The trajectory spends most of the time in the broad potential well of the nonconvecting state. The fact that the well is broad however favours the trajectory wandering across the convection threshold y_0 and entering the convecting regime. We call this *wandering monostability*: deterministically there is only one stable state, but the coarse-grained probability n_c for a convecting year is significantly larger than P_c and may even exceed 0.5. Point W in Fig. 4.12b denotes the domain of wandering monostability, and panel f displays how one tail of the pdf peak stretches into the convecting regime. Note that the pdf of any nonlinear potential well model driven by Gaussian noise will easily extend into unstable regimes; the effect of the wandering monostability is a consequence of the physically relevant convection threshold.

Wandering monostability in the stochastic 1S model is an important effect because the coarse-grained probability n_c is well motivated by the physics of deep convection. In the real ocean like in the 1S model, the convection events are quite short, but still achieve the vertical mixing of the water column. The forcing of the upper layer has to bring the vertical density gradient just beyond the point of neutral stratification to start a convection event. In the deterministic 2TS model it is the seasonal cycle

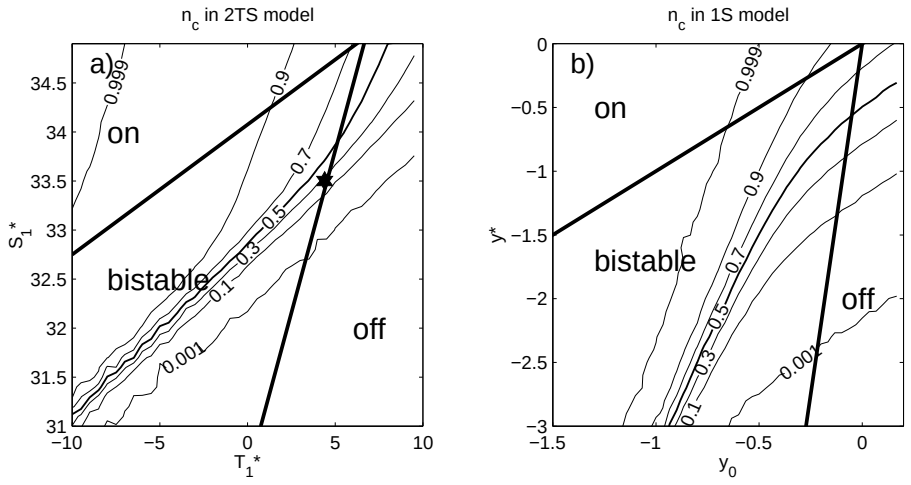

Figure 4.13: *Comparison of the n_c values for (a) the 2TS model and (b) the 1S model. Parameter values are for 1S the tutorial ones, and for 2TS the optimal parameter set, where T_1^* and S_1^* vary, and $\sigma = 18°C$. The asterisk in (a) denotes the position of the optimal parameters as given in table 2.2.*

that drives the model over the convection threshold and back again; in the stochastic 2TS model convection events may be additionally triggered (or suppressed) by the stochastic forcing; and in the 1S model the stochastic forcing alone causes the convection events.

4.7.3 Stability in the 1S and in the 2TS model

The n_c stability diagram from the 1S model compares well with the n_c stability diagram from the 2TS model (Fig. 4.13). The probabilities n_c are directly comparable because they were defined in the same way: given a model run, n_c is the probability for a convecting year, or the fraction of years with a convection event out of the total number of years. (See section 3.4 for details on n_c in the 2TS model). Both stability diagrams show the areas of effective monostability in the flanks of the bistable domain, and both stability diagrams show wandering monostability in parts of the "off" domain. The different curvature of the contours may stem from the 2TS model not having a simple potential well structure. The noise has two opposite effects: deterministic bistability is turned into effective monostability, and due to the coarse-grained probability n_c areas of deterministic monostability show wandering monostability.

The focus is here on the satisfying qualitative accordance of the two stochastic stability diagrams. Fig. 4.13 demonstrates that the simplifications introduced in this chapter – reducing the number of variables from four to

one and compensating for the neglected seasonal cycle with stronger stochastic forcing and the coarse-grained statistics – have not destroyed the structure of the model dynamics. Therefore, the very simple and theoretically biased 1S model could be used to gain an extended stochastic-dynamical understanding of the stochastic 2TS model, the stochastic climate model of deep convection.

The two different pictures are crucial, too, for the interpretation of the position of the estimated parameter set (the asterisk in Fig. 4.13a). In the deterministic picture, a shift of the parameters to lower S_1^* or larger T_1^* looks quite dramatic due to the associated loss of stability for the "on" state. In the stochastic picture, however, the same parameter changes only lead to a smaller probability for a convecting year, but obviously no abrupt threshold crossing will happen. In other words, the presence of variability in the climate system has a moderating influence here, since the *stochastic* model can stay temporarily in a state that is *deterministically* unstable.

4.8 Conclusions

The aim of this chapter is to gain a theoretical understanding of the stochastic stability and variability properties of the 2TS model of deep convection. As a tool, a simplified "1S" box model with one single variable is developed. Its output is analyzed by the probability for convection to occur during any given year. The dependence of this probability on the model parameters can be explained in principle analytically. The stochastic stability diagrams of both the 1S and the 2TS model are qualitatively rather similar. In the domain of wandering monostability, only the nonconvecting state is stable, but there are frequent excursions over the threshold into the convecting regime. The stochastic stability features, too, are described analytically.

The 1S model draws a very simple picture of a potentially convective water column. The nonconvecting state is associated with a broad and shallow potential well. The convecting state, in contrast, is modeled as a narrow and deep potential well. The two wells are connected at the convection threshold, where the vertical density gradient is zero. The model's variable, the vertical salinity gradient, can be imagined as an over-damped ball moving in these wells. A crossing of the threshold is associated with a transition from one model regime to the other. For instance, if convection is absent in the beginning, a strong salinity anomaly may reduce the density gradient until convection starts. Such a forcing pushes the ball away from the broad well over the threshold into the deep well of the convecting state. The stochastic forcing continuously provides anomalies of all kinds that keep pushing the ball around.

Now, the model parameters may be such that a stable convecting state does not exist. The respective potential well is replaced by a mere upward

sloping potential curve. The ball cannot stay for long on this slope, it will roll back into the nonconvecting state. Yet, it is a crucial feature that it can still cross the threshold to convect temporarily. The salinity anomaly mentioned above may be strong enough to start convection, but then it is diluted by the mean freshwater flux until convection ceases.

The idea of such transient excursions into the convecting regime is at the core of this chapter and provides the answer to the third guiding question (GQ3). Through the stochastic forcing these excursions occur frequently, even if the convecting state does not exist. This effect is called wandering monostability. Although unstable from a deterministic perspective, the deep water formation process continues because the model trajectory wanders into the convecting regime. In this picture, the positive salinity feedback is represented by the ball's damped motion in the broad nonconvecting potential well.

Another phenomenon of stochastic stability is dubbed effective monostability (see also Monahan (2002)). Here the stochastic forcing has the effect that one of two existing potential wells is almost never visited. Both effects, effective monostability and wandering monostability, highlight the need to take climate variability into account when analyzing the stability of climate states.

Many nonlinearities and thresholds arise from the climate system's high complexity. In their vicinity, the presence of variability may lead to clear qualitative stability changes. The deterministic picture of one or more distinct stable climate states is replaced by the stochastic picture of jumps between different regimes. It is then this wandering which has to be taken as the overall climate state, and not only the averages of the observed quantities. The example of flickering convection as seen in CGCMs (P. Wu, pers. comm.) supports this picture.

The simplicity of the 1S model allows us to see analytically how the probability for an unusually long residence in one regime depends on the model parameters. The mean residence times depend exponentially on parameters like the mean freshwater forcing, or its variance. The probability for the residence times to exceed a given limit grows exponentially as well, as Khatiwala et al. (2001) already concluded. However, this holds only for small values of that probability. Our results show that the initially quick growth slows down to a linear increase if that probability has intermediate values.

We expect that our results are valid for other physical systems with a relevant threshold. For instance, simple stochastic models have been successfully applied to convection in the tropical atmosphere (Lin and Neelin, 2000; Yano et al., 2001), and they have also been suggested as convection parameterizations in the atmosphere (Palmer, 2001). Threshold-crossing statistics have been used to analyze the time series of the NAO (Wunsch, 1999) and the 100 kyr cycle found in paleoclimatic records (Wunsch, 2002).

Large-scale atmospheric variability is seen as a wandering between different regimes (Corti et al., 1999). Our stochastic stability concepts, in particular the wandering monostability, might help to understand these climatic processes as well.

The benefits of a simple model are bought at the price of drawbacks. For the stochastic climate model ansatz, it is sometimes not easy to justify the necessary time scale separation. As in other studies our results depend on how that ansatz is carried out in detail. However, the qualitative agreement between the 1S and the 2TS model showed that the basic stability and variability properties persist if the time scale separation border is shifted. Having compared the 2TS model with the observational data increases the credibility of our results further. Nevertheless it needs to be studied more deeply to what extent the concepts of stochastic stability developed here capture the stability properties of GCMs and of the real deep convection sites. Some strategies to do so are suggested in the next, concluding chapter.

Chapter 5

Summary and outlook

The aim of this thesis is to study the stability and variability of open-ocean deep convection. We have developed a conceptual modeling approach to address the guiding questions we put forward in the Introduction (section 1.1). This involves a two-box, or 2TS model, as well as a more simplified one-box, or 1S model. With the box models we explore how the slow heat and freshwater fluxes, caused by exchange with the atmosphere and horizontal eddy mixing, interact with the vigorous vertical convective mixing in the water column. In contrast to previous studies on the stability of deep convection, our box model study starts from observational data and not from simulations with general circulation models (GCMs). The 2TS model is successful in reproducing the observed variability of deep convection in the Labrador Sea. The 1S model is a tool to understand this variability analytically. Both models are based on the main hypothesis (section 1.2) of a locally active positive salinity feedback.

In the following we summarize what our answers to the guiding questions are and what support for the main hypothesis was found, and we outline what our results imply for future research.

5.1 Answers to the guiding questions

How sensitive is deep convection in the Labrador Sea to changes in climate? The answer to this question depends on whether a deterministic or a stochastic approach is used.

The deterministic approach focuses on the mean state in the model. The variability of the forcing is accounted for only in the form of single anomalies. In this approach, the results from fitting the 2TS model to the Labrador Sea data suggest that the Labrador Sea is bistable: deep convection occurs either regularly every winter, or not at all. A shift of the *mean* external forcing to either warmer or fresher conditions may easily drive the model out of the bistable domain into a domain where only the nonconvecting state is stable.

Deep convection then stops abruptly, meaning that it is highly sensitive to climate change. Under constant mean forcing, a short *anomaly* in the external forcing – for instance, a warm winter – can still switch convection off. But due to the bistability, a following anomaly may switch it on again.

The stochastic approach includes the variability of the external forcing. The heat flux variability caused by the weather is represented by an additional stochastic term in the 2TS model. This variability leads to frequent switches between the two model states. If now the *mean* of the forcing is changed, then the reaction is not an abrupt end of convection. Rather, the frequency of convection events decreases. The *variance* of the forcing has a similar effect. Hence both the mean and the variance can compete. If there is, say, less convection due to a surface freshening, this may be compensated for by increased variability in the heat fluxes. The stochastic approach is more realistic since it includes the observed atmospheric variability. Furthermore, frequent convection switches are seen in observational data (e.g. Dickson et al. (2002)) as well as in simulations of high-resolution coupled GCMs (P. Wu, pers. comm.).

Both approaches, the deterministic and the stochastic, indicate a high sensitivity to changes in the forcing. The climate changes expected for the next decades might significantly reduce the frequency of deep convection events in the Labrador Sea. From CGCM simulations it was concluded that even a total shutdown has to be envisaged (Wood et al., 1999; Schweckendiek and Willebrand, 2002). In any case, the regional climate will presumably be seriously affected.

Defining "stability" as "sensitivity to perturbations" — what role does external climate variability play in determining the stability of deep convection? The two different approaches followed to answer the first question show that the variability of the external forcing has a fundamental impact on the stability of deep convection. With the frequent jumps into and out of the model states, we can define stochastic stability by the mean residence time in either state. This adds a quantitative aspect to the concept of stability that enables us to compare how stable the two states are. The stochastic 2TS model shows that in the Labrador Sea the nonconvecting state is clearly more stable than the convecting one.

Moreover, due to the stochastic forcing the model trajectory accesses deterministically unstable regimes: convection events occur even if the convecting state is unstable. This explains why the convection frequency depends smoothly on the forcing parameters, rather than showing abrupt changes.

A consequence of the positive salinity feedback together with the external variability is a small probability for long nonconvecting phases. Such phases have been observed in the Greenland Sea (Rhein, 1996). Our results suggest that they are not necessarily signs of a complete convection shutdown; they might be part of the natural variability instead. A sufficiently strong surface

forcing (like an exceptionally cold winter) might well trigger further deep convection events.

How can this role be explained theoretically in a simple conceptual way? In the 1S model the nonconvecting state is imagined as a broad potential well. A ball is moving in this well, pushed by the stochastic forcing and damped by the mean freshwater input. Somewhere up one slope there is a threshold where the convecting regime starts. If the 1S model time series is analyzed in a newly developed coarse-grained way, we can see that for the theoretical understanding it is not essential how the ball is pushed beyond the threshold, or how convection is started: by the stochastic forcing, by the seasonal cycle, or both. This explains why a change in the amount of variability may counteract a change in the mean forcing in the effect on the frequency of deep convection events.

If in the convecting regime there is no second potential well, then convection is not a stable state, but can still occur. We call this effect wandering monostability, and we explain it analytically using the coarse-grained statistics. The analytical results demonstrate that wandering monostability is closely linked to short mean residence times in the nonconvecting state.

The self-sustaining effect of deep convection is, in this picture, explained by the long time scale for the external freshwater fluxes. Saline surface waters that stem from a deep convection event are only gradually diluted and are able to favour deep convection in the following winter.

5.2 Support for the main hypothesis

With the main hypothesis we have proposed a mechanism by which a nonconvecting phase begins and ends. To switch deep convection off, all it takes is a short anomaly in the forcing that suppresses deep convection in one winter. The mean freshwater input then accumulates in the upper layer. This strengthens the vertical density gradient. Due to this positive salinity feedback it becomes increasingly harder to restart convection, but eventually a strong cold (or salt) anomaly might be sufficient to achieve this.

Previous studies (Dickson et al., 1996; Lilly et al., 1999) followed the idea that a freshwater forcing anomaly is permanently present during a nonconvecting phase, and that the positive salinity feedback is not relevant to maintain it. Our results show that the assumption of an active positive salinity feedback is equally in accordance with the observational data. The 2TS model that is built on our main hypothesis successfully represents these data. It explains the upper layer salinity drift in the absence of deep convection, and it reproduces some features of the observed variability. The 1S model still contains the same mechanism, giving qualitatively the same results. All these issues render the main hypothesis quite plausible, estab-

lishing an alternative to the idea of Dickson et al. (1996) and Lilly et al. (1999).

In addition, there is some evidence from other studies that supports our hypothesis. Houghton and Visbeck (2002) very recently employed the positive salinity feedback in order to explain the observed salinity variability in the Labrador Sea. Moreover, there is a number of modelling studies demonstrating how forcing anomalies can switch deep convection on and off (Lenderink and Haarsma, 1994, 1996; Rahmstorf, 1995c), and showing that the positive salinity feedback is indeed operating at deep convection sites (Rahmstorf, 1994; Voss and Mikolajewicz, 2001).

5.3 Implications and outlook

Our conceptual picture obviously works well for the Labrador Sea. Ongoing work on modelling Labrador Sea deep convection with more refined conceptual models (F. Straneo, pers. comm.) indicates that qualitative changes of our picture are not to be expected. This confirms what we have seen from the similarity of the results from the 2TS and the 1S models.

We suspect that our results apply to some extent to the other important deep convection site in the North Atlantic, the Greenland Sea. The situation there however differs from the Labrador Sea. First, the presence of sea-ice cannot be neglected. Then, it is rather the intermediate than the deep waters that contribute to the THC overturning. Confined by bottom topography, the deep waters form a reservoir of very dense water that is replenished by rare deep convection events. All this suggests at least a three-box conceptual model; with variable box depths the relevant difference between shallow and deep convection would be better represented.

The IPCC recently assessed that climate change can appear as a shift in a mean value, or as a change in the preference of different regimes (Houghton et al., 2001). A shifting mean value is easily explained in a purely deterministic picture. Our stochastic models show how changes in regime preferences (in terms of the mean residence time) can come about. In this broader context wandering monostability means that certain climate regimes might still be visited when they have become unstable before – but they are visited less often. Such a lowered frequency may seem less critical than an abrupt climate change. Yet, this lowering may cause problems when, for instance, one deals with atmospheric regimes that decide about drought or precipitation for large populated regions.

With the 1S model we show that, under a surface freshening or warming, the mean length of nonconvecting phases increases exponentially. However, the probability of these phases to be longer than a given length shows an exponential growth only initially, when this probability is small, but then shifts to a linear increase. Since the simplicity of the 1S model allows to

generalize these results, they are an extension to the study of Khatiwala et al. (2001). For instance, if the probability for a long drought period in a certain region increases with global warming, then the results from our study suggest that this increase is exponential only when the absolute value of this probability is small.

Our results bear some implications for ocean modelling with GCMs. There are studies that analyze the deep water formation processes in a purely deterministic way: the OGCM is forced with an atmosphere climatology, and attains a mean circulation state after the spinup (e.g. Marotzke and Scott (1999)). This corresponds to the deterministic approach we started with. Our results emphasize that the inclusion of variability might give a different picture of the deep water formation process and its stability.

Of course, there are also many model studies that include a fully variable, dynamic atmosphere. Such coupled GCMs give a detailed picture of its impact on ocean dynamics (e.g. Timmermann et al. (1998); Cooper and Gordon (2002)). However, their high computational cost prevents extensive sensitivity studies, and the delicate balance between the atmosphere and the ocean modules may lead to unrealistic instabilities when performing perturbation experiments.

Stochastically forced OGCMs can fill the gap between coupled GCMs and conceptual models. Provided that the stochastic atmosphere is not too complex, it is still possible to carry out long model runs that allow a sound statistical analysis. For instance, thermohaline oscillations (Skagseth and Mork, 1998) and the variability of the thermohaline circulation (Holland et al., 2001) have been studied in this way. There are still many uncertainties about the appropriate parameterization of deep convection and the associated deep water formation. Yet, some OGCMs represent it acceptably, and in approximately the right locations (e.g. CLIO, see Goosse (1998)). By forcing such an OGCM stochastically, the sensitivity of specific convection sites could be studied explicitly. One would then have a three-step modeling approach, consisting of the box model, the stochastically driven OGCM, and the full coupled GCM. Such an approach will certainly deepen the conceptual understanding of the stability of deep convection.

Bibliography

Aeberhardt, M., Blatter, M., and Stocker, T. F. (2000). Variability on the century time scale and regime changes in a stochastically forced zonally averaged ocean-atmosphere model. *Geophys. Res. Lett.*, 27:1303–1306.

Arnold, L. (2001). Hasselmann's program revisited: the analysis of stochasticity in deterministic climate models. In Imkeller, P. and von Storch, J.-S., editors, *Stochastic Climate Models*, volume 49 of *Progress in Probability*, pages 141–158. Birkhäuser, Boston.

Belkin, I. M., Levitus, S., Antonov, J., and Malmberg, S.-A. (1998). Great Salinity Anomalies in the North Atlantic. *Prog. Oceanog.*, 41:1–68.

Böning, C. W., Bryan, F. O., Holland, W. R., and Döscher, R. (1996). Deep-water formation and meridional overturning in a high-resolution model of the North Atlantic. *J. Phys. Oceanogr.*, 26:1142–1164.

Cessi, P. (1994). A simple box model of stochastically forced thermohaline flow. *J. Phys. Oceanogr.*, 24:1911–1920.

Cessi, P. (1996a). Convective adjustment and thermohaline excitability. *J. Phys. Oceanogr.*, 26:481–491.

Cessi, P. (1996b). Grid-scale instability of convective-adjustment schemes. *J. Mar. Res.*, 54:407–420.

Cessi, P. and Young, W. R. (1996). Some unexpected consequences of the interaction between convective adjustment and horizontal diffusion. *Physica D*, 98:287–300.

Cooper, C. and Gordon, C. (2002). North Atlantic oceanic decadal variability in the Hadley Centre coupled model. *J. Clim.*, 15:45–72.

Corti, S., Molteni, F., and Palmer, T. N. (1999). Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, 398:799–802.

Delworth, T., Manabe, S., and Stouffer, R. J. (1993). Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model. *J. Clim.*, 6:1993–2011.

Dickson, R. R., Lazier, J., Meincke, J., Rhines, P., and Swift, J. (1996). Long-term coordinated changes in the convective activity of the North Atlantic. *Prog. Oceanog.*, 38:241–295.

Dickson, R. R., Meincke, J., Malmberg, S.-A., and Lee, A. J. (1988). The "Great Salinity Anomaly" in the Northern North Atlantic 1968–1972. *Prog. Oceanog.*, 20:103–151.

Dickson, R. R., Yashayaev, I., Meincke, J., Turrell, B., Dye, S., and Holfort, J. (2002). Rapid freshening of the deep North Atlantic Ocean over the past four decades. *Nature*, 416:832–837.

Einstein, A. (1905). Über die von der molekular-kinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, F(17):549–560.

Flato, G. M. and Boer, G. J. (2001). Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, 28(1):195–198.

Fofonoff, N. P. and Millard Jr., R. C. (1984). Algorithms for computation of fundamental properties of sea water. Unesco Tech. Pap. Mar. Sci 44, UNESCO, Paris.

Frankignoul, C. and Hasselmann, K. (1977). Stochastic climate models. Part II: Application to sea surface temperature anomalies and thermocline variability. *Tellus*, 29:289–305.

Freidlin, M. I. and Wentzell, A. D. (1998). *Random Perturbations of Dynamical Systems*. Springer, New York, 2nd edition.

Ganachaud, A. and Wunsch, C. (2000). Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, 408:453–456.

Ganopolski, A. and Rahmstorf, S. (2002). Abrupt glacial climate changes due to stochastic resonance. *Phys. Rev. Lett.*, 88(3):038501.

Gardiner, C. W. (2002). *Handbook of stochastic methods for physics, chemistry and the natural sciences*, volume 13 of *Springer series in Synergetics*. Springer, Berlin, 2nd edition.

Gill, A. E. (1982). *Atmosphere-Ocean Dynamics*. Academic Press, San Diego (USA).

Goosse, H. (1998). *Modelling the Large-Scale Behaviour of the Coupled Ocean-Sea Ice System*. PhD thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Griffies, S. M. and Tziperman, E. (1995). A linear thermohaline oscillator driven by stochastic atmospheric forcing. *J. Climate*, 8:2440–2453.

Häkkinen, S. (1999). A simulation of thermohaline effects of a Great Salinity Anomaly. *J. Clim.*, 12:1781–1795.

Hall, A. and Manabe, S. (1997). Can local linear stochastic theory explain sea surface temperature and salinity variability? *Clim. Dyn.*, 13:167–180.

Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, 62(2):251–342.

Hasselmann, K. (1976). Stochastic climate models, Part I: Theory. *Tellus*, 28:473–485.

Hillaire-Marcel, C., de Vernal, A., Bilodeau, G., and Weaver, A. J. (2001). Absence of deep-water formation in the Labrador Sea during the last interglacial period. *Nature*, 410:1073–1077.

Hirschi, J., Sander, J., and Stocker, T. F. (1999). Intermittent convection, mixed boundary conditions and the stability of the thermohaline circulation. *Clim. Dyn.*, 15:277–291.

Holland, M. M., Bitz, C. M., Eby, M., and Weaver, A. J. (2001). The role of ice-ocean interactions in the variability of the North Atlantic thermohaline circulation. *J. Clim.*, 14(5):656–675.

Houghton, J. T., Ding, Y., Griggs, D., Noguer, M., van der Linden, P. J., and Xiaosu, D., editors (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*. Cambridge University Press, Cambridge.

Houghton, R. W. and Visbeck, M. H. (2002). Quasi-decadal salinity fluctuations in the Labrador Sea. *J. Phys. Oceanogr.*, 32(2):687–701.

Kalnay, E., Kanamitsu, M., Kitler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebiszusaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-years reanalysis project. *Bull. Amer. Meteor. Soc.*, 77:437–471.

Khatiwala, S., Schlosser, P., and Visbeck, M. (2002). Rates and mechanisms of water mass transformations in the Labrador Sea as inferred from tracer observations. *J. Phys. Oceanogr.*, 32(2):666–686.

Khatiwala, S., Shaw, B. E., and Cane, M. A. (2001). Enhanced sensitivity of persistent events to weak forcing in dynamical and stochastic systems: Implications for climate change. *Geophys. Res. Lett.*, 28:2633–2636.

Khatiwala, S. and Visbeck, M. (2000). An estimate of the eddy-induced circulation of the Labrador Sea. *Geophys. Res. Lett.*, 27:2277–2280.

Khatiwala, S. P., Fairbanks, R. G., and Houghton, R. W. (1999). Freshwater sources to the coastal ocean off Northeastern North America: evidence from $H_2^{18}O/H_2^{16}O$. *J. Geophys. Res.*, 104:18241–18255.

Klinger, B. A., Marshall, J., and Send, U. (1996). Representation of convective plumes by vertical adjustment. *J. Geophys. Res.*, 101(C8):18,175–18,182.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics*. Springer-Verlag, Berlin.

Knutti, R. and Stocker, T. F. (2002). Limited predictability of the future thermohaline circulation close to an instability threshold. *J. Clim.*, 15(2):179–186.

Kramers, H. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304.

Kuhlbrodt, T., Titz, S., Feudel, U., and Rahmstorf, S. (2001). A simple model of seasonal open ocean convection. Part II: Labrador Sea stability and stochastic forcing. *Ocean Dynamics*, 52(1):36–49.

Kunze, E. and Sandford, T. B. (1996). Abyssal mixing: Where it is not. *J. Phys. Oceanogr.*, 26:2286–2296.

Lab Sea Group (1998). The Labrador Sea deep convection experiment. *Bull. Amer. Meteorol. Soc.*, 79:2033–2058.

Lazier, J. R. N. (1973). The renewal of Labrador Sea Water. *Deep-Sea Research*, 20:341–353.

Lazier, J. R. N. (1980). Oceanographic conditions at Ocean Weather Ship Bravo, 1964–1974. *Atmosphere-Ocean*, 18:227–238.

Lazier, J. R. N. (1988). Temperature and salinity changes in the deep Labrador Sea, 1962–1986. *Deep-Sea Research*, 35(8):1247–1253.

Leadbetter, M., Lindgren, G., and Rootzen, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York, Heidelberg, Berlin.

Lenderink, G. and Haarsma, R. J. (1994). Variability and multiple equilibria of the thermohaline circulation associated with deep-water formation. *J. Phys. Oceanogr.*, 24:1480–1493.

Lenderink, G. and Haarsma, R. J. (1996). Modeling convective transitions in the presence of sea ice. *J. Phys. Oceanogr.*, 36(8):1448–1467.

Levitus, S. (1982). Climatological atlas of the world ocean. Technical Report NTIS PB83-184093, NOAA/ERL GFDL, Princeton, N.J.

Lilly, J. M., Rhines, P. B., Visbeck, M., Davis, R., Lazier, J. R. N., Schott, F., and Farmer, D. (1999). Observing deep convection in the Labrador Sea during winter 1994–1995. *J. Phys. Oceanogr.*, 29:2065–2098.

Lin, J. W.-B. and Neelin, J. D. (2000). Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, 27(22):3691–3694.

Lind, P. G., Titz, S., Kuhlbrodt, T., Corte-Real, J. A., Kurths, J., Gallas, J. A. C., and Feudel, U. (2002). Coupled bistable maps: A tool to study convection parametrization in ocean models. *Physical Review E*, submitted.

Lohmann, G. and Schneider, J. (1999). Dynamics and predictability of Stommel's box model. A phase-space perspective with implications for decadal climate variability. *Tellus*, 51A:326–336.

Marotzke, J. (1991). Influence of convective adjustment on the stability of the thermohaline circulation. *J. Phys. Oceanogr.*, 21:903–907.

Marotzke, J. and Scott, J. R. (1999). Convective mixing and the thermohaline circulation. *J. Phys. Oceanogr.*, 29:2962–2970.

Marshall, J. and Schott, F. (1999). Open-ocean convection: Observations, theory, and models. *Rev. Geophys.*, 37:1–64.

Mertens, C. (2000). *Open-ocean convection in the Labrador and Greenland Seas: plume scales and interannual variability*. PhD thesis, Christian-Albrechts-Universität Kiel.

Molemaker, M. J. and Dijkstra, H. A. (2000). Stability of a cold core eddy in the presence of convection: hydrostatic versus nonhydrostatic modeling. *J. Phys. Oceanogr.*, 30:475–494.

Monahan, A. H. (2002). Stabilisation by noise of climate regimes in a simple model: implications for stability of the thermohaline circulation. *J. Phys. Oceanogr.*, 32:2072–2085.

Monahan, A. H., Timmermann, A., and Lohmann, G. (2002). Comments on "Noise-induced transitions in a simplified model of the thermohaline circulation". *J. Phys. Oceanogr.*, 32(3):1112–1116.

Munk, W. and Wunsch, C. (1998). Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res. I*, 45:1977–2010.

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

Palmer, T. N. (1999). A nonlinear dynamical perspective on climate prediction. *J. Clim.*, 12:575–591.

Palmer, T. N. (2001). A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic-dynamic parameterisation in weather and climate prediction models. *Quart. J. Roy. Meteorol. Soc.*, 127(572):279–304.

Pierce, D. W., Barnett, T. P., and Mikolajewicz, U. (1995). Competing roles of heat and freshwater flux in forcing thermohaline oscillations. *J. Phys. Oceanogr.*, 25:2046–2064.

Rahmstorf, S. (1993). A fast and complete convection scheme for ocean models. *Ocean Modelling*, 101:9–11.

Rahmstorf, S. (1994). Rapid climate transitions in a coupled ocean-atmosphere model. *Nature*, 372:82–85.

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

Rahmstorf, S. (1995b). Climate drift in an ocean model coupled to a simple, perfectly matched atmosphere. *Clim. Dyn.*, 11:447–458.

Rahmstorf, S. (1995c). Multiple convection patterns and thermohaline flow in an idealized OGCM. *J. Clim.*, 8:3028–3039.

Rahmstorf, S. (2001). A simple model of seasonal open ocean convection. Part I: Theory. *Ocean Dynamics*, 52(1):26–35.

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

Rhein, M. (1996). Convection in the Greenland Sea. *J. Geophys. Res.*, 101(C8):18183–18192.

Roemmich, D. H. and Wunsch, C. (1985). Two transatlantic sections: Meridional circulation and heat flux in the subtropical North Atlantic Ocean. *Deep-Sea Res.*, 32:619–664.

Sathiyamoorthy, S. and Moore, G. W. K. (2002). Buoyancy flux at Ocean Weather Station Bravo. *J. Phys. Oceanogr.*, 32(2):458–474.

Schweckendiek, U. and Willebrand, J. (2002). Global warming in a regional model of the Atlantic Ocean – ECHAM4/OPYC3 in FLAME 4/3. Poster at EGS XXVII General Assembly.

Send, U. and Marshall, J. (1995). Integral effects of deep convection. *J. Phys. Oceanogr.*, 25:855–872.

Skagseth, Ø. and Mork, K. A. (1998). Stability of the thermohaline circulation to noisy surface buoyancy forcing for the present and a warm climate in an ocean general circulation model. *J. Phys. Oceanogr.*, 28:842–857.

Smith, S. D. and Dobson, F. W. (1984). The heat budget at Ocean Weather Station Bravo. *Atmosphere-Ocean*, 22:1–22.

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

Stommel, H. (1987). *A view of the sea*. Princeton University Press, Princeton.

Stommel, H. M. and Young, W. R. (1993). The average T-S relation of a stochastically forced box model. *J. Phys. Oceanogr.*, 23:151–158.

Tett, S. F. B., Johns, T. C., and Mitchell, J. F. B. (1997). Global and regional variability in a coupled AOGCM. *Clim. Dyn.*, 13:303–323.

Timmermann, A., Latif, M., Voss, R., and Grötzner, A. (1998). Northern hemispheric interdecadal variability: a coupled air-sea mode. *J. Clim.*, 11:1906–1931.

Timmermann, A. and Lohmann, G. (2000). Noise-induced transitions in a simplified model of the thermohaline circulation. *J. Phys. Oceanogr.*, 30:1891–1900.

Titz, S. (2002). *Bifurcations of oceanic overturning and convection in conceptual models of the thermohaline circulation*. PhD thesis, Universität Potsdam, Potsdam, Germany.

Vellinga, M. (1998). Multiple equilibria in ocean models as a side effect of convective adjustment. *J. Phys. Oceanogr.*, 28:621–633.

Visbeck, M., Marshall, J., Haine, T., and Spall, M. (1997). Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.*, 27:381–402.

Visbeck, M. and Rhein, M. (2000). Is bottom boundary layer mixing slowly ventilating Greenland Sea deep water? *J. Phys. Oceanogr.*, 30(1):215–224.

von Storch, H. and Zwiers, F. (1999). *Statistical analysis in climate research*. Cambridge University Press, Cambridge.

Voss, R. and Mikolajewicz, U. (2001). Long-term climate changes due to increased CO_2 concentration in the coupled atmosphere-ocean general circulation model ECHAM3/LSG. *Clim. Dyn.*, 17:45–60.

Weisse, R., Mikolajewicz, U., and Maier-Reimer, E. (1994). Decadal variability of the North Atlantic in an ocean general circulation model. *J. Geophys. Res.*, 99:12411–12421.

Welander, P. (1982). A simple heat-salt oscillator. *Dyn. Atmos. Oceans*, 6:233–242.

Wood, R. A., Keen, A. B., Mitchell, J. F. B., and Gregory, J. M. (1999). Changing spatial structure of the thermohaline circulation in response to atmospheric CO_2 forcing in a climate model. *Nature*, 399:572–575.

Wunsch, C. (1999). The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteorol. Soc.*, 80(2):245–255.

Wunsch, C. (2002). The spectral description of climate change including the 100ky energy. *Clim. Dyn.*, in press.

Yano, J. I., Fraedrich, K., and Blender, R. (2001). Tropical convective variability as $1/f$ noise. *J. Clim.*, 14(17):3608–3616.

Acknowledgements

I am very grateful to many people for the support and advice they have given me, each of them in his own way. I could not have completed this thesis without their help. Here are some of them:

Eva Bauer
Ulrike Feudel
Alexa Griesel
Christian Großheim
Hermann Held
Brigitte Kausch
Björn Kuhlbrodt
Dietrich Kuhlbrodt
Axel Kunellis
Jürgen Kurths
Adam Monahan
Marisa Montoya
Miguel Angel Morales Maqueda
Stefan Rahmstorf
Julia Saal
Matthias Schneider
Udo Schwarz
Sven Titz
Alexei Zaikin