Nonlinear dynamics of atmospheric regime transitions

Niet-lineaire dynamica van atmosferische regime-overgangen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof. dr. W. H. Gispen, ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op vrijdag 17 januari 2003 des middags te 12.45 uur

door

Daniël Theodoor Crommelin

geboren op 10 augustus 1972 te Delft

Promotoren: Prof. dr. ir. J. D. Opsteegh
              faculteit der Natuur- en Sterrenkunde, Universiteit Utrecht;
              Koninklijk Nederlands Meteorologisch Instituut
              Prof. dr. F. Verhulst
              faculteit der Wiskunde en Informatica, Universiteit Utrecht

Daher ist die Aufgabe nicht sowohl, zu sehn was noch Keiner gesehn hat, als, bei Dem, was Jeder sieht, zu denken, was noch Keiner gedacht hat.
A. Schopenhauer

Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe sind blind.
I. Kant

Contents

Introduction

Some think a scientists job is done once the laws governing a system have been written down. Others have a different opinion. Among them are those dealing with fluid dynamics. Their work started, rather than ended, when in 1822 Claude-Louis Navier published his version of the natural law that describes the temporal evolution of a fluid or gas. Now known as the Navier-Stokes equations to acknowledge the contribution of George Stokes, the laws of fluid motion have a misleadingly simple appearance, in sharp contrast with the variety and complexity of the phenomena described by them. The Navier-Stokes equations are nonlinear, implying all kinds of feedback effects that are hard to understand. The nonlinearity is responsible for the situation that, although the laws of fluid motion have been known now for almost two centuries, many major questions concerning systems governed by these laws are still unanswered.

I.1   Atmospheric low-frequency dynamics: regime behaviour

This thesis deals with a notorious example of a fluid dynamical system: the atmosphere of planet earth. The main focus will be on the low-frequency behaviour of the extratropical planetary-scale atmospheric flow, i.e. on the behaviour of flow characteristics of over 5000 km. of spatial extent, outside the tropics, on timescales of a week or longer. A major role in that behaviour is played by so-called *regimes*: preferred patterns of the planetary-scale atmospheric circulation. Regimes can capture the atmospheric flow for some time (typically one or two weeks), after which the flow makes a transition to another regime. The planetary-scale flow pattern is more or less frozen when the atmosphere is caught in a regime. It sets the large-scale boundary conditions for the advection (horizontal transport) of smaller scale (so-called synoptic scale) atmospheric phenomena, such as depressions, that on their turn determine day-to-day weather conditions.

Atmospheric Low-Frequency Variability (LFV) is the name for variations in the atmosphere on timescales of a week or longer (see Haines (1994) for a review). Variations on timescales beyond several months are gathered under the name Ultra-Low Frequency Variability (ULFV). The switching back and forth between regimes dominates the variability of the midlatitude (i.e. non-tropical) planetary-scale flow at the intraseasonal timescale (weeks to months). The central role of regimes in atmospheric LFV is reflected in the teleconnection patterns, correlation patterns relating geographical regions all across the northern hemisphere (Wallace and Gutzler, 1981). The two most prominent telecon-

nection patterns, the North Atlantic Oscillation (NAO) pattern and the Pacific/North American (PNA) pattern, are also found in the regimes. The standing wave oscillations assumed to underlie these teleconnections can be interpreted as transitions between different regimes.

The notion of atmospheric circulation regimes found its first articulation in studies like Baur (1947), Namias (1950) and Rex (1950a,b). Based on the experience of synoptic meteorologists, it was recognised that the atmospheric flow over the North-Atlantic possessed two states in which it could be caught for some time: a *zonal* state of strong west-to-east flow, and a *blocked* state characterised by weak zonal flow and large wave amplitudes, resulting in more meridionally (north-south) oriented flow. The alternation, or vacillation, between those two states has been called the *index cycle*. Later on, the notion of zonal and blocked flow over the Atlantic evolved into the concept of regimes that are of hemispheric extent. For the study of regimes, meteorological intuition was replaced by statistical techniques, such as probability density estimation used on observational datasets. Since Benzi et al. (1986) and Sutera (1986), this approach has been increasingly successful in identifying preferred states of the large-scale Northern-Hemisphere (NH) flow. Moreover, the results of different regime studies have become more and more similar to each other, see e.g. Cheng and Wallace (1993), Kimoto and Ghil (1993a), Smyth et al. (1999) and Monahan et al. (2001).

An important step towards a better understanding of the dynamics that drive regime behaviour and LFV was made by Charney and DeVore (1979). They suggested that regimes correspond with equilibrium solutions (steady states, fixed points) of the equations describing the evolution of the large-scale flow. Their study made clear that in a severely truncated model of midlatitude flow over topography ("mountains"), multiple equilibria can exist that represent different flow regimes. The existence of more than one equilibrium is due to the presence of topography in the model and the nonlinearity of the model equations. This multiple equilibria theory triggered and inspired numerous subsequent studies. In particular, a connection was made with the mathematical theory of dynamical systems. The recognition that a finite-dimensional set of ordinary differential equations, embodying an atmosphere model, can be viewed as a dynamical system opened up a whole new class of concepts and techniques of considerable mathematical sophistication. In particular the technique of (numerical) bifurcation analysis turned out to be useful in the search for a better understanding of atmospheric flow dynamics, and has been used by Vickroy and Dutton (1979), Källén (1981, 1982), Legras and Ghil (1985), Yoden (1985a,b), De Swart (1988a,b, 1989), Itoh and Kimoto (1996, 1999) and many others.

I.2   Meteorology and dynamical systems

The connection between dynamical systems theory and meteorology starts, some time before the emergence of the multiple equilibria theory, with the much celebrated work by Lorenz. His milestone 1963 article *Deterministic nonperiodic flow*, on a highly simplified model for convection in the atmosphere, showed that the difference between two model simulations starting from almost identical initial states grew very rapidly. Also, the model did not evolve towards an equilibrium state or towards a (quasi-)periodic solution,

although it was entirely deterministic. It took several years before these results, published in a journal on meteorology, got through to the mathematics community, but eventually they turned out to connect with mathematical ideas on nonlinear systems, as developed by e.g. Smale (1963) and Shil'nikov (1965). The extreme initial state sensitivity, and the nonperiodicity of the bounded asymptotic solution became the hallmarks of systems that behave, in a mathematical sense, chaotic. From the many other results and models put forward by Lorenz, the model presented in Lorenz (1984) is in particular worth mentioning. It describes baroclinic instability of a zonal flow in a highly simplified atmospheric context, and has been analysed thoroughly from a dynamical systems perspective by Shil'nikov et al. (1995) and Van Veen (2002a,b). Broer et al. (2002) couple the model to a periodic forcing, representing seasonal variations of the solar forcing, and analyse it in great detail.

In meteorology, the initial state sensitivity (also referred to as exponential error growth) has important consequences for weather prediction. The inevitably limited knowledge of the exact state of the atmosphere, needed to specify an initial state for weather prediction models, means that the model simulation and the real atmosphere will evolve differently after some time. This limits the prediction of weather: the predictive power of model results, even at the high level of resolution and sophistication of present day models, is close to zero for predictions more than, roughly, ten days ahead. This fact has been the center of much research effort over the years. However, the focus of this thesis lies elsewhere. The presumed chaotic nature of atmospheric flow, and its implications for predictability, is not the main theme here. The central issue will be regime behaviour and atmospheric LFV.

Although a lot of information has been gathered on atmospheric LFV, the understanding of the dynamical structures and mechanisms that play a role in the generation of LFV is still limited. For the development of such understanding, the theory of dynamical systems can be of great use, as it allows to single out specific dynamical mechanisms and study them in detail. The availability of advanced numerical analysis techniques is particularly helpful in this respect. It is known that the low-frequency behaviour of the atmosphere is dominated by a rather small number of variability patterns (see e.g. Kimoto and Ghil, 1993a). Moreover, these patterns are almost equivalent barotropic (Wallace and Blackmon, 1983), which means that they have no vertical variations in the direction of the horizontal air flow. As a consequence, the role of baroclinic processes is possibly of secondary importance in the low-frequency behaviour. Baroclinic instabilities may be merely a trigger for low-frequency variability, without playing a major role in determining the structure of the low-frequency dynamics. The small number of variability patterns that dominate atmospheric LFV and the role of regimes give the strong suggestion that it must be a rather simple dynamical structure, consisting of e.g. only a few steady states and periodic orbits, that generates the core of atmospheric low-frequency dynamics. The analysis of atmospheric LFV from a dynamical systems perspective is the most natural way to uncover this structure.

I.3 Regime transitions

The application of dynamical systems theory on models of atmospheric flow has provided a lot of information on the structure of (model) circulation regimes. However, the phenomenon of regime *transitions* has so far not been satisfactorily explained. If, à la Charney and DeVore, regimes are associated with stable fixed points, or with other simple attracting invariant sets such as periodic orbits or tori, one encounters the problem that a system will not leave an attracting solution whereas the real atmosphere continously hops from one regime to another. Therefore, a stochasticity assumption is often invoked to explain regime transitions. Noise, representing the effect of physics and dynamics (e.g. baroclinic processes) not included in a model, can kick that model out of the basin of attraction of one regime and into another. This has been studied, by adding stochastic perturbations to a low-order model, in e.g. Egger (1981), Benzi et al. (1984), De Swart and Grasman (1987) and Sura (2002). However, the addition of noise is not necessary to trigger regime transitions. Even in deterministic low-order models transitions can occur. Examples are the model of De Swart (1988), with 10 variables, and that of Legras and Ghil (1985) with 25 variables. Even the model of Charney and DeVore can be brought to regime transitions without noise when the zonal forcing profile is taken more general than in the original Charney and DeVore study, as will become clear in chapter three of this thesis. It is therefore worthwhile to consider how transitions can be generated by the internal, deterministic dynamics of a system, in the absence of noise.

From statistical studies of model data as well as observational data, it has been known for some time that transitions between regimes are not an entirely random process. In Mo and Ghil (1987, 1988), Molteni et al. (1990) and Kimoto and Ghil (1993b), up to seven regimes/clusters are identified in various datasets (both from observations and models). By counting the transitions between the various regimes, these studies made clear that not only preferred flow regimes exist, but also preferred transitions between (some of) the regimes. Plaut and Vautard (1994), studying the interplay between low-frequency oscillations and regimes, also found that transitions are not purely random.

Itoh and Kimoto (1996, 1997, 1999) propose chaotic itinerancy as an explanation for the preferences and inhomogeneities in regime transitions. Using multi-layer quasi-geostrophic models of moderate complexity (L2T15 and L5T21), they detect multiple attractors that are identified as regimes. By changing a parameter (horizontal diffusivity or static stability), these attractors lose their stability one by one, thereby admitting regime transitions. When all attractors have lost their stability, the model shows irregular transitions between the remnants of the former attractors, and thus between the regimes. This behaviour is called chaotic itinerancy. The preferred order of the transitions is related to the order in which the attractors lost their stability. This interesting result once more emphasizes the inhomogeneities in regime transitions, which must be due to the nonlinear nature of the system under study. However, it does not provide much insight in the dynamics that determine the transitions, as the loss of stability of the various attractors is not analysed. It remains unknown what dynamical processes cause and drive the transitions.

I.4   Heteroclinic connections

The deterministic view leads to the interpretation of regime transitions as manifestations of *heteroclinic connections*. Heteroclinic connections (or orbits) are solutions of (deterministic) model equations that connect one invariant set, e.g. a steady state solution, with another. Homoclinic orbits connect an invariant set with itself. Hetero- resp. homoclinic bifurcations are bifurcations in which such connections are created. They are mentioned a few times in the literature as a possible mechanism for transitions between regimes. Legras and Ghil (1985) use a 25-dimensional equivalent-barotropic model, perform a bifurcation analysis of its steady states and find, by numerical integration, transitions between regimes. Heteroclinic orbits as a possible explanation are mentioned, but not expanded on. Kimoto and Ghil (1993b) and Weeks et al. (1997) also suggest the possibility that regime transitions are related to heteroclinic connections. However, no studies have been performed yet which focus on the role of heteroclinic connections in atmospheric regime behaviour. To my knowledge, this thesis is the first study which explicitly adresses this issue.

A set of fixed points and heteroclinic connections forming a closed loop is called a *heteroclinic cycle*. Such a cycle would correspond to a cycle of regime transitions. In model studies, the existence of a heteroclinic cycle is usually related to the presence of symmetry in the model. In the most general situation, several parameters of a model must be tuned in order to create a closed cycle of heteroclinic connections. The number of parameters that must be tuned is called the codimension. The lower the codimension of a phenomenon, the greater the chance to encounter that phenomenon in practical situations. When symmetries are present in a model, they can reduce the codimension of a heteroclinic cycle. Symmetries can even make cycles generic (i.e. no parameters need to be tuned). In such a case, a cycle will persist under perturbations that respect the symmetries of the model equations. The cycle is then said to be robust. A review of the theory on robust heteroclinic cycles is given in Krupa (1997).

The source of symmetry in the large-scale atmospheric flow system is not a priori clear. One could argue that the real atmosphere possesses no symmetry and that the role of a heteroclinic cycle in atmospheric dynamics is therefore questionable. However, that would be premature. Even when a cycle is broken its remnants may still influence the system dynamics. Therefore, the behaviour of a system with perturbed symmetry can still show signs of a heteroclinic cycle. In fact, since regime behaviour is irregular, the atmosphere system cannot be exactly on a heteroclinic cycle – some kind of perturbation must be present. One can think of several near-symmetries in the atmosphere system. The topography could be considered as a perturbation of the axisymmetry. In the northern hemisphere, where the symmetry-breaking due to topography is more severe than in the southern hemisphere, the topography has an important wavenumber two component, implying a (nearby) $\mathbb{Z}_2$ symmetry. Moreover, the laws governing the atmosphere can posses symmetries not yet uncovered. See e.g. Blender and Névir (1991), who find a new symmetry when studying the Lie-symmetries of the quasi-geostrophic potential vorticity equation on a β-plane without topography.

In the fluid dynamics community, the possibility of heteroclinic connections as an explanation for transitions between steady states has been recognised before. Examples were

found in e.g. Rayleigh-Benard convection (Proctor and Jones, 1988) and the turbulent boundary layer (Aubry et al. (1988), Armbruster et al. (1988), Holmes et al. (1996,1997)). Knobloch and Moehlis (2000) provide an overview of mechanisms that produce bursting in hydrodynamical systems (sudden transitions from and to regular, repeatedly realized states); again, heteroclinic connections play an important role. They interpret transitions between different types of behaviour as a form of bursting. Although no examples are mentioned from geophysical fluid dynamics, such a concept of bursting could equally well apply to the case of regime-transitions in atmospheric dynamics. The methods and concepts used in the study of these fluid-dynamical examples could be very useful in the investigation of regime behaviour and low-frequency variability of the atmosphere.

I.5   What lies ahead

The four chapters following this introduction reflect a conceptual development stretching several years. Insights gained in later papers were, obviously, not yet available when starting the research leading up to this thesis. As a result, chapter one is not yet focused on regime transitions but rather on a general view of the dynamics driving a simple atmosphere model. Previous studies of low-order atmosphere models investigated specific mechanisms playing a role in atmospheric dynamics, such as topographic or baroclinic instability. The low-order model considered in chapter one tries to capture the behaviour of the global atmosphere, covering the two hemispheres including the tropics. The interaction between two (slowly) oscillating modes, both with mainly Northern Hemispheric variability, was found to be able to reproduce the main features of the model dynamics: growth and decay of turbulent kinetic energy. This behaviour was related to the presence of a homoclinic orbit of double-focus type, nearby in parameter space. The orbit, connecting a steady state of mainly zonal circulation with itself, generates chaotic behaviour which can explain the nonzero variability on ultra-long timescales. A brief inspection of the behaviour of a moderately complex, barotropic model of Northern Hemisphere flow showed similar behaviour.

A closer investigation of the same hemispheric model in chapter two resulted in a more subtle picture of the dynamics. The Probability Density Function (PDF) of the two most dominant patterns of variability showed two preferred states, or regimes. A numerical calculation of steady states made clear that these regimes could be associated with the dynamically most important steady states of the model equations. Heteroclinic connections between these steady states were approximated numerically using a new algorithm. The algorithm is based on adjoint modelling techniques which stem from 4D-VAR data assimilation in weather forecasting. The approximated (remnants of) connections match well with statistically found phase space preferences of the model during regime transitions.

The hypothesis that remnants of heteroclinic connections guide regime transitions, inspired the model investigation in chapter two, but was itself left without firm mathematical support. In chapter three such support is searched by analysing a low-order model of atmospheric flow at midlatitudes. Regime transitions have been observed in different barotropic models, in which only topographic and barotropic instabilities are

present. Therefore, the simplest model possible that combines these two instability mechanisms is used in chapter three. Such a model is provided by De Swart (1988a,b, 1989), a six-variable model that is essentially the same as the model of Charney and DeVore (1979), also studied by Yoden (1985), except for a more general zonal forcing profile. By parameter tuning, the Hopf bifurcation corresponding to barotropic instability can be made to coincide with one of the saddle-node bifurcations that are due to the topography in the model. This coincidence is called a fold-Hopf bifurcation. In the unfolding of the the truncated normal form of the fold-Hopf bifuraction, a heteroclinic cycle is present that connects two steady states associated with different regimes. In the full model, this cycle is perturbed, leaving orbits homoclinic to one of the equilibria.

Having studied the deterministic aspects of regime transitions in models, in the final chapter the attention is shifted to the real atmosphere. The behaviour of the real planetary-scale atmospheric flow is driven both by noise, stemming mainly from synoptic-scale, high-frequency atmospheric processes, and by the internal nonlinear dynamics of the planetary-scale flow system. An attempt is made to isolate the latter influence from the former, in an analysis of observational data of the atmosphere system during winter. The data are projected onto the unit sphere in a reduced 3-dimensional phase space, spanned by the leading three preferred variability patterns (Empirical Orthogonal Functions, or EOFs; see section 1.2 for a short explanation or Preisendorfer (1988) for a detailed account). The probability density function (PDF) associated with the data on the unit sphere shows clear maxima that correspond with flow regimes also found in previous studies. Concepts from the theory of Markov chains and phase space partitions are used to bring out the non-diffusive, reversible dynamics of the large-scale atmospheric flow. It is found that the role of such dynamics is small but becomes increasingly important in low-pass filtered data, i.e. in the low-frequency behaviour of the atmosphere. By inspecting the asymmetries of transitions between different phase space cells, a preferred path over the unit sphere of the reduced phase space is detected. This path connects regimes of zonal and blocked flow in the Atlantic sector. The analysis of observational data thus confirms the point of view, adopted throughout this thesis, that deterministic dynamical processes play a role in the low-frequency dynamics of the planetary-scale atmospheric flow, in particular in the regime behaviour.

The results presented in this thesis have been submitted for publication:

Chapter one: D.T. Crommelin, *Homoclinic dynamics: a scenario for atmospheric ultra-low-frequency variability*, J. Atmos. Sci. 59 (2002), 1533-1549.

Chapter two: D.T. Crommelin, *Regime transitions and heteroclinic connections in a barotropic atmosphere*, J. Atmos. Sci, accepted for publication.

Chapter three: D.T. Crommelin, J.D. Opsteegh, F. Verhulst, *A mechanism for atmospheric regime behaviour*, J. Atmos. Sci, submitted for publication.

Chapter four: D.T. Crommelin, *Observed non-diffusive dynamics in large-scale atmospheric flow*, J. Atmos. Sci, submitted for publication.

Chapter 1

Homoclinic dynamics: a scenario for atmospheric ultra-low-frequency variability

Abstract. In this paper, a link will be established between atmospheric ultra-low-frequency variability (ULFV) and the occurrence of homoclinic dynamics in models of large-scale atmospheric flow. It is known that uncoupled atmosphere models possess significant variability on very long timescales (years to decades), which must be generated by internal atmospheric dynamics. The mathematical structure of this long-timescale variability is investigated, using a global two-layer atmosphere model formulated in terms of preferred flow patterns (EOFs). Due to its efficient formulation, this model can simulate an atmospheric flow with realistic features, using only a small number of degrees of freedom. The 10-dimensional version of the model possesses both nonzero ultra-low-frequency variability and several realistic short timescales. The essence of the ultra-long-timescale behaviour of the 10D model, which manifests itself as bursting in the atmospheric turbulent energy, can be represented by a 4-dimensional subsystem. In this subsystem, strong evidence for the existence of a homoclinic orbit is found. The chaotic dynamics generated by the homoclinic orbit explains the occurrence of ultra-long timescales in the model. It is argued that hints of homoclinic dynamics can also be found in more complex models. As an example, a T21 barotropic atmosphere model (231 dimensions) of the northern hemisphere is shown to possess behaviour that suggests the existence of a homoclinic orbit.

1.1 Introduction

Atmospheric variability on very long timescales, purely generated by its own internal dynamics, is not a very well understood feature of the atmosphere. The short timescales (up to a week) are the classical subject of meteorological research, since they are within the typical reach of weather prediction. Changes in the atmosphere on timescales from a week to a few months, often identified as "Low-Frequency Variability" or LFV, have been studied rather intensely over the last decades, using and studying concepts such as blocking, multiple equilibria and teleconnection patterns. In contrast, atmospheric behaviour on timescales beyond several months (also called Ultra-Low-Frequency Variability or ULFV)

has hardly received any attention apart from the context of coupled models: atmospheric ULFV is usually associated with other, "slower" components of the climate system (e.g. the oceans) and is therefore mostly studied in models in which the atmosphere is coupled to the oceans, to sea-ice, etc. The ULFV of the atmosphere in itself, of its internal dynamics has been ignored for a long time.

The limited number of papers that have been devoted to the ULFV generated by the internal dynamics of realistic atmosphere models mostly have a descriptive/phenomenological content: using models of varying complexity, powerspectra of the frequency domain are calculated, showing nonzero variability on very long timescales (years and longer). James and James (1989) use a 5-level, T21 primitive equation model with both constant and periodic forcing to arrive at powerspectra most of which show peaks at decadal periods over a red-noise type background. James and James (1992), using the same model, find maximum variability in a broad frequency-band, roughly between 5 and 40 years. They also investigate the spatial structure of the ULFV in their model, using EOFs of the zonal-mean zonal flow. From this information, they hypothesize that the interactions between zonal flow and baroclinic waves induce the long timescale behaviour. This hypothesis is further investigated (and supported) in James et. al. (1994). Pielke and Zeng (1994) run the Lorenz 1984 model for 1100 years and find nonzero spectral power for periods longer than a year. Kurgansky et al. (1996) use a low-order (8 components) hemispheric, baroclinic model and find maximum variability between 3 and 44 years from a 1100 years integration. Dethloff et. al. (1998) run a 20-dimensional baroclinic model with orography for 10 000 years, thus obtaining results with better statistical reliability than in previous studies. They find significant variability at ultra-low frequencies, with a maximum at a period of about 10 years. Handorf et. al. (1999), as part of their study of a coupled atmosphere-ocean model, run a moderately complex atmosphere model with prescribed, seasonally varying sea surface temperatures for 1000 years and find modes that are strongly peaked at 9 years.

The above studies follow roughly the same strategy: a relatively simple atmospheric model, which allows for computationally cheap and fast integrations of considerable length, is integrated for a very long time (up to 10 000 years). The data thus obtained is used to calculate powerspectra of various quantities, such as spectral components. They all show significant spectral power on very long timescales, orders of magnitudes longer than the typical response timescale of the atmosphere.

In this paper, another approach is followed. Rather than aiming at more detailed information about the precise spatio-temporal structure of the atmospheres ultra-low-frequency variability, we want to focus on the mathematical structure and mechanism(s) underlying the ultra-long timescale behaviour of atmospheric models. Using techniques and concepts from dynamical systems theory we shall try to gain more insight into the generation of ULFV in the atmosphere. This goal puts us in the middle of the gap between mathematics/dynamical systems on the one hand and meteorology/climatology on the other hand. In order to be analysable for mathematicians, nonlinear systems must be low-dimensional; however, in order to be somewhat realistic, atmosphere models cannot be that low-dimensional. The famous Lorenz models of 1963 and 1984, for instance, are often thought of as metaphors of the real atmosphere. However, even though these models are inspired by meteorological problems, they are not even remotely capable (nor meant)

to describe global atmospheric dynamics. The (many) results that are known by now on the creation of chaos and strange attractors in these models are not necessarily valid for global atmospheric models, and the insights the Lorenz models provide should therefore only be applied with great care to the global atmosphere. The mechanisms and sources of chaos and long timescales can be totally different in global models.

In order to combine a small number of variables with an acceptable degree of realism, we have turned towards the efficient model of Achatz and Branstator (1999): an atmospheric model formulated in terms of Empirical Orthogonal Functions (EOFs), closed with an empirical closure scheme. From a General Circulation Model (GCM) dataset (NCAR CCM version 0B) EOFs were calculated and used as basis functions for a two-layer model based on filtered (quasi-geostrophic-, or QG-like) equations. After truncation of this model, an empirical closure scheme was applied in which forcing and linear terms of the model equations were changed in order to minimize tendency errors (errors in the time derivative) produced by the model. With as few as 30 degrees of freedom, the resulting model has a remarkably good performance when it comes to climate simulation. We shall work with the 10-dimensional version of this model, as it is a low-dimensional version which still shows both nonzero variability on ultra-long timescales and several realistic short timescales. The 10D model will be analysed in order to find out what mechanism(s) and source(s) create chaos and long timescales (in other words: ultra-low-frequency variability). Due to the degree of realism of the 10D model, and its derivation from a GCM, the results obtained on the creation of chaos and ULFV are more likely to represent realistic atmospheric characteristics than the results on the Lorenz models (or any other model analysis we know of). To our knowledge, no attempt has been made before to gain more insight into the creation of chaos in an atmosphere model that is designed to be global and realistic, rather than to be merely a metaphor.

This paper is organised as follows. In section 2, the derivation and formulation of the EOF-model is briefly described. In section 3 we compare the 10 EOF model with the 30 EOF model and the GCM, in order to get an impression of its degree of realism. Section 4 contains an analysis of the 10 EOF model. We will show that the essence of the long timescale behaviour can be represented by a 4-dimensional subsystem of the 10D system; this is discussed in section 5, which also contains results of the analysis of the 4D subsystem. In section 6 we will discuss to what extent the obtained results and insights can be expected to be valid for more complex models, such as the 30 EOF model and the GCM. In section 7, the output of an entirely different model, a T21 standard barotropic model, will be used to further support the relevance of the results for more complex models. Section 8 concludes the paper with a summary and outlook.

1.2   Model formulation

The atmosphere model we wish to analyse has to meet two conflicting requirements: the model must be low-dimensional, and it must be realistic. Although it is impossible to resolve this conflict entirely, one can at least chose to work with a model that uses its few degrees of freedom as efficiently as possible, thereby providing an acceptable compromise. Such efficiency can be created by describing atmospheric flow in terms of preferred flow

patterns rather than in terms of spherical harmonics. The model formulated by Achatz and Branstator (1999, AB hereafter) employs *Empirical Orthogonal Functions* (EOFs) as its basis functions. Such a basis captures a maximum amount of atmospheric variability with a limited number of basis functions. In combination with an empirical closure scheme, a highly efficient model is obtained. This model will be the subject of our investigation; its derivation will be described briefly in this section. For more detailed information on the derivation the reader is referred to AB.

The recipe used by AB in order to make their model efficient consists of three steps. First, streamfunction EOFs are calculated from GCM data. Then the expansion coefficients of these EOFs are used as the variables of a two-layer filtered (QG-like) model. After truncating this EOF-based model, an empirical closure scheme is applied in which the linear and constant terms of the filtered equations are modified such that the resulting set of equations minimizes the tendency error for states in the GCM dataset.

The GCM data used are two 50 000 day integrations (half daily output), under perpetual January conditions, performed with version 0B of the NCAR Community Climate Model (an R15 model with nine levels in the vertical). Timeseries describing the streamfunction at 200 and 700 mb are extracted from the data; the atmospheric state at time t is now represented by a 990-dimensional vector $\mathbf{X}(t)$, which contains the spectral coefficients of the streamfunction at both 200 and 700 mb. In the 990-dimensional state space in which $\mathbf{X}(t)$ lives, the vector which describes the direction of the largest variance of the data points (using a total energy metric M to calculate "distance" in state space) is identified as the first EOF. The second EOF is the vector which describes the direction of the second largest variance and which is orthogonal to the first EOF. And so on: the EOFs are all orthogonal to each other and are arranged, in descending order, according to the amount of variance (in this case total energy variance) explained by each of them.

EOFs can be computed by solving the eigenvalue problem

$$\mathsf{C}\,\mathsf{M}\,\mathbf{p}^i = \lambda_i^2\,\mathbf{p}^i \tag{1.1}$$

in which $\mathbf{p}^i$ is the EOF with index i; the state vector $\mathbf{X}$ and $\mathbf{p}^i$ live in the same vector space: $\mathbf{X}, \mathbf{p}^i \in \mathbf{R}^{990}$. M is the total energy metric, the total energy of a state $\mathbf{X}$ is given by $\mathbf{X}^{\mathrm{T}}\,\mathsf{M}\,\mathbf{X}$. λ_i^2 is the fraction of the variance explained by $\mathbf{p}^i$. Finally, C is the lag zero covariance matrix (overbars denote time averages):

$$\mathsf{C} = \overline{(\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X}^{\mathrm{T}} - \overline{\mathbf{X}^{\mathrm{T}}})} \tag{1.2}$$

In summary: the EOFs are calulated from the timeseries $\{\mathbf{X}(t)\}$ using a total energy norm, yielding a set of patterns $\{\mathbf{p}^i\}$, $i = 1, \ldots, 990$. Each EOF corresponds to a streamfunction anomaly pattern at both 200 and 700 mb. The EOF spectrum is rather flat: about 500 EOFs are needed to explain 90% of the total energy variance.

The two-layer filtered model used by AB was derived by Lorenz (1960); it is essentially the same as a classical quasi-geostrophic model. Projection of the (discretized) model equations onto the EOFs, i.e. replacing the state vector $\mathbf{X}(t)$ by $\overline{\mathbf{X}} + \sum_i a_i(t)\,\mathbf{p}^i$, yields an equation for $\mathbf{a}$, the 990-dimensional vector containing the EOF expansion coefficients a_i:

$$\dot{\mathbf{a}} = \mathbf{F}^a + \mathsf{L}^a\,\mathbf{a} + \mathsf{N}(\mathbf{a}, \mathbf{a}) \tag{1.3}$$

with time-independent forcing $\mathbf{F}^a$, linear operator L^a and quadratic nonlinearities $\mathsf{N}(\mathbf{a}, \mathbf{a})$. This system is truncated, leaving an m-dimensional set of ODEs prescribing the time-evolution of $\mathbf{a} = (a_1, \ldots, a_m)^{\mathrm{T}}$

Finally, $\mathbf{F}^a$ and L^a are replaced by new operators $\mathbf{F}$ and L, which were determined such that the resulting set of equations produces tendencies $\dot{\mathbf{a}}$ which resemble the corresponding tendencies in the GCM data as closely as possible. This empirical correction of $\mathbf{F}$ and L accounts for the effect of unresolved physics and dynamics on the explicitly described patterns, e.g. the effect of the EOFs that were lost in the truncation, or dynamical processes in the tropics which are not captured by the QG equations (thereby making the model more realistic than ordinary QG models). The final result is an m-dimensional set of ODEs:

$$\dot{\mathbf{a}} = \mathbf{F} + \mathsf{L}\,\mathbf{a} + \mathsf{N}(\mathbf{a}, \mathbf{a}) \tag{1.4}$$

Note that (1.4) models the time-evolution of *deviations* of the atmosphere from the GCM mean state $\overline{\mathbf{X}}$. Anomalous, not total flow is described; the "background" flow $\overline{\mathbf{X}}$ is given and constant. The atmospheric mean state resulting from the model (1.4) can be different from the mean state in the GCM, so $\sum_{i=1}^{m} \bar{a}_i\,\mathbf{p}^i$ can be nonzero.

A last remark on the energy in the model: the total turbulent energy (the phase space distance to $\mathbf{a} = 0$ using the total energy norm) is equal to the sum of squares of the EOF coefficients, i.e. $E = \mathbf{a} \cdot \mathbf{a}$ (Euclidean inner product). This is due to the use of the total energy norm in the calculation of the EOFs. E is unaffected by the nonlinear terms, which means that $\mathbf{a} \cdot \mathsf{N}(\mathbf{a}, \mathbf{a}) = 0$. The nonlinear interactions redistribute energy but do not change the overall energy in the system (this is a common property of fluid dynamical models – the advection terms conserve energy).

1.3   The 10D model

The performance of the empirical EOF model was tested in various ways by AB, e.g. by inspecting the climatology of the 500 EOF model, by comparison with an optimized standard spectral model and by calculation of tendency errors for EOF model versions of varying dimension (using the GCM data as reference). The last test showed that the relative tendency error was quite large (0.9 - 0.4) but monotonically decreasing for increasing number of dimensions. In contrast, the relative first moment error and relative second moment error showed minima ($\mathcal{O}(10^{-2})$) in the range between roughly 30 and 70 dimensions. For higher dimensions, these error measures increased. Apparently, quite simple EOF models (e.g. 30-dimensional), although not describing the instantaneous tendency very well, can capture a lot of the dynamics on longer timescales (reflected in the low first and second moment errors). This feature makes these models suitable for studying (ultra-)low-frequency variability of the atmosphere.

The analysis in this paper will focus on the 10-dimensional EOF model. The relative tendency error and the relative first moment error of the 10D model are comparable to those of the 30D model; its relative second moment error is substantially larger. The total energy variance explained by the first 10 EOFs is 21.7 % ; 30 EOFs explain 34.9 % . Integrations of the 10D and 30D models and inspection of the timeseries of the turbulent energy resulting from these integrations shows (figure 1.1) that both models exhibit large
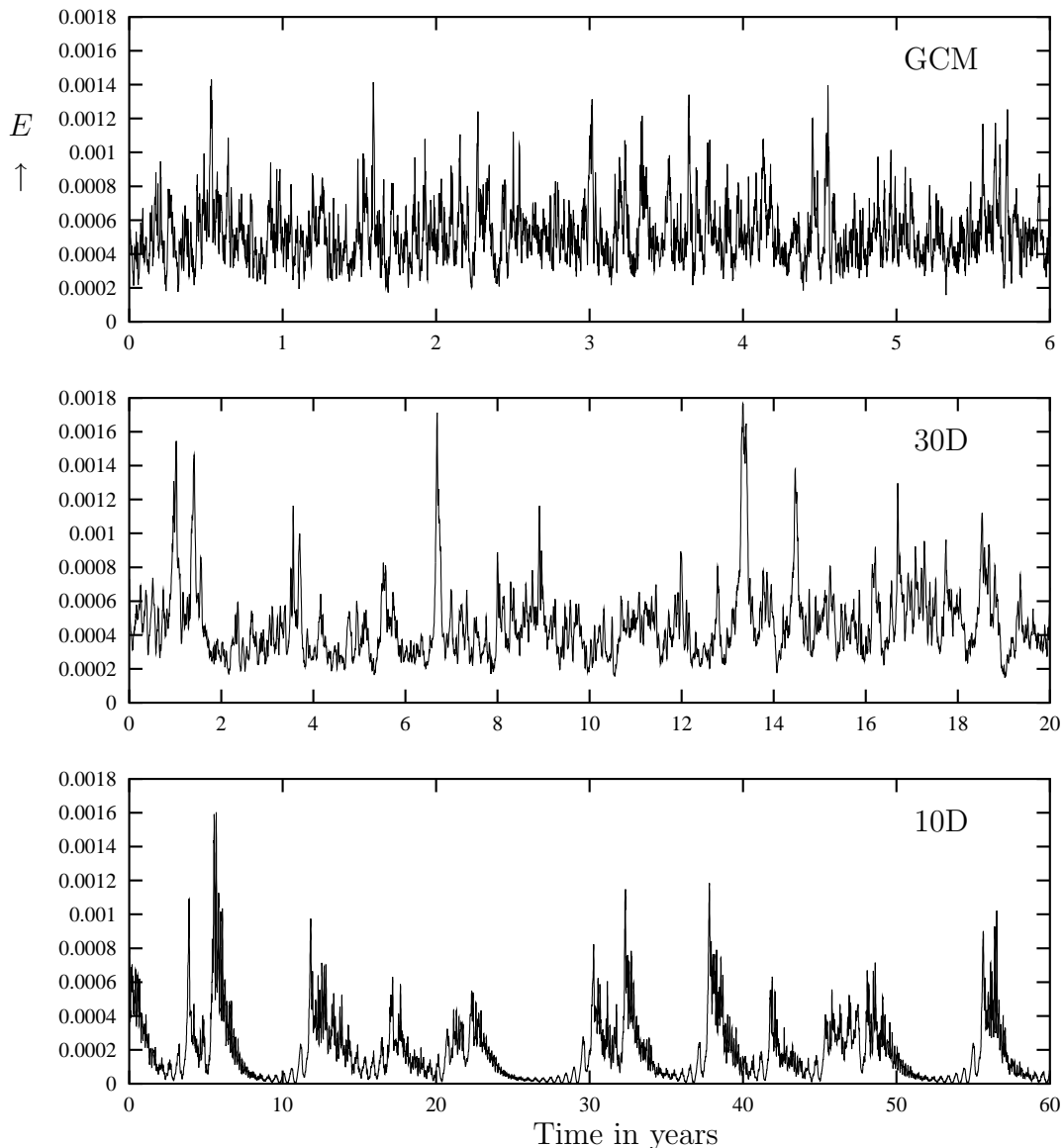
Figure 1.1: *Timeseries of turbulent energy in GCM, 30D and 10D models (note the different timescales).*

variations in energy: periods of relative quietness (low energy) and sudden outbursts of energy during which the atmosphere is highly active. The same phenomenon can be seen in the GCM data if we restrict ourselves to the large spatial scales, i.e. to the leading EOFs. In figure (1.1) the turbulent energy of the GCM data projected onto the first 30 EOFs is shown. The bursting-like behaviour is the dominant aspect of the long timescale behaviour of all three models. The energy-bursts become further apart in time and more pronounced, compared to the GCM, in the 30D model and especially in the 10D model. One could say that the strong, characteristic bursts in the 10D model are schematized, overpronounced versions of the bursts in the 30D model and the GCM. Their heights, however, are of the same magnitude.

Powerspectra of the turbulent energy of the 30D and 10D models, calculated from

Figure 1.2: *Powerspectra of turbulent energy in 30D and 10D models*

integrations of 3×10^6 days length (8×10^3 years), are provided in figure 1.2. No turbulent energy spectrum of the GCM data is shown, d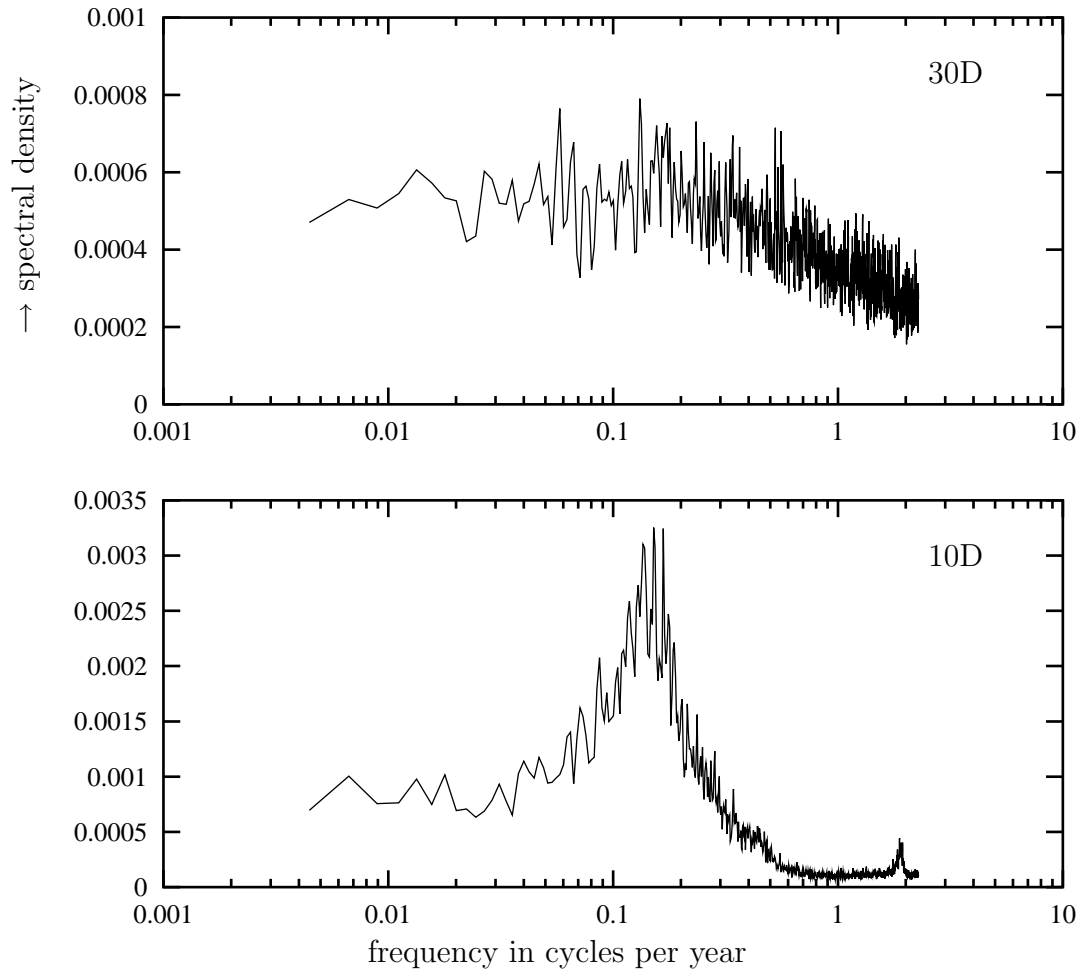ue to the limited time span of that data. The spectrum of the 10D model clearly shows nonzero variability at very low frequencies. It also shows a strong peak at a frequency of about 0.15 cycles per year (i.e. a period of about 6.7 years); this peak is due to the bursting behaviour of the model. The 30D model spectrum has the same characteristics: nonzero ULFV and a maximum at periods slightly less than 10 years (although this maximum is much less pronounced than in the 10D model). Apparently, the strong 6.7 year peak in the 10D model is much too pronounced but not unrealistic, since its frequency matches both with the 30D model and with peaks in the frequency spectra of previous studies: Handorf et. al. (1999) find a peak of 9 years, Dethloff et. al. (1998) find a decadal peak, and both James and James (1989, 1992) and Kurgansky et. al. (1996) find maximal variability at decadal periods.

The 10D model not only possesses nontrivial long timescale behaviour, it also reproduces some essential features of the GCM and the 30D model on shorter timescales. In figure 1.3 powerspectra of the EOF1-coefficient are shown for the GCM, the 30D EOF model and the 10D model. The complex spectrum stemming from the GCM gets more and more simplified when going to the simpler models. The spectrum gets 'stripped down'

to its most essential features: peaks at periods of about 20, 30, 65 and 190 days (and nonzero variability at ultra-low frequencies). In fact, the spectrum of the 30D model shows 5 dominant "short" (< 1 yr) periods. To reproduce these in a simpler model one needs at least 10 dimensions. The 10D model under consideration indeed reproduces those 5 periods: the above-mentioned 4, and a fifth (of 14 days) which isn't very well visible in the EOF1 spectrum but which is nevertheless present in the model (as will become clear in the next section).

The above observations suggest the validity of using the 10D EOF model for studying the mechanisms of atmospheric ULFV. It is simple and yet describes global atmospheric dynamics; it has some degree of realism due to its derivation from a GCM. The model possesses "short" timescales that are also prominent in the 30D model and the GCM; on long timescales it shows nonzero variability associated to bursting behaviour also present
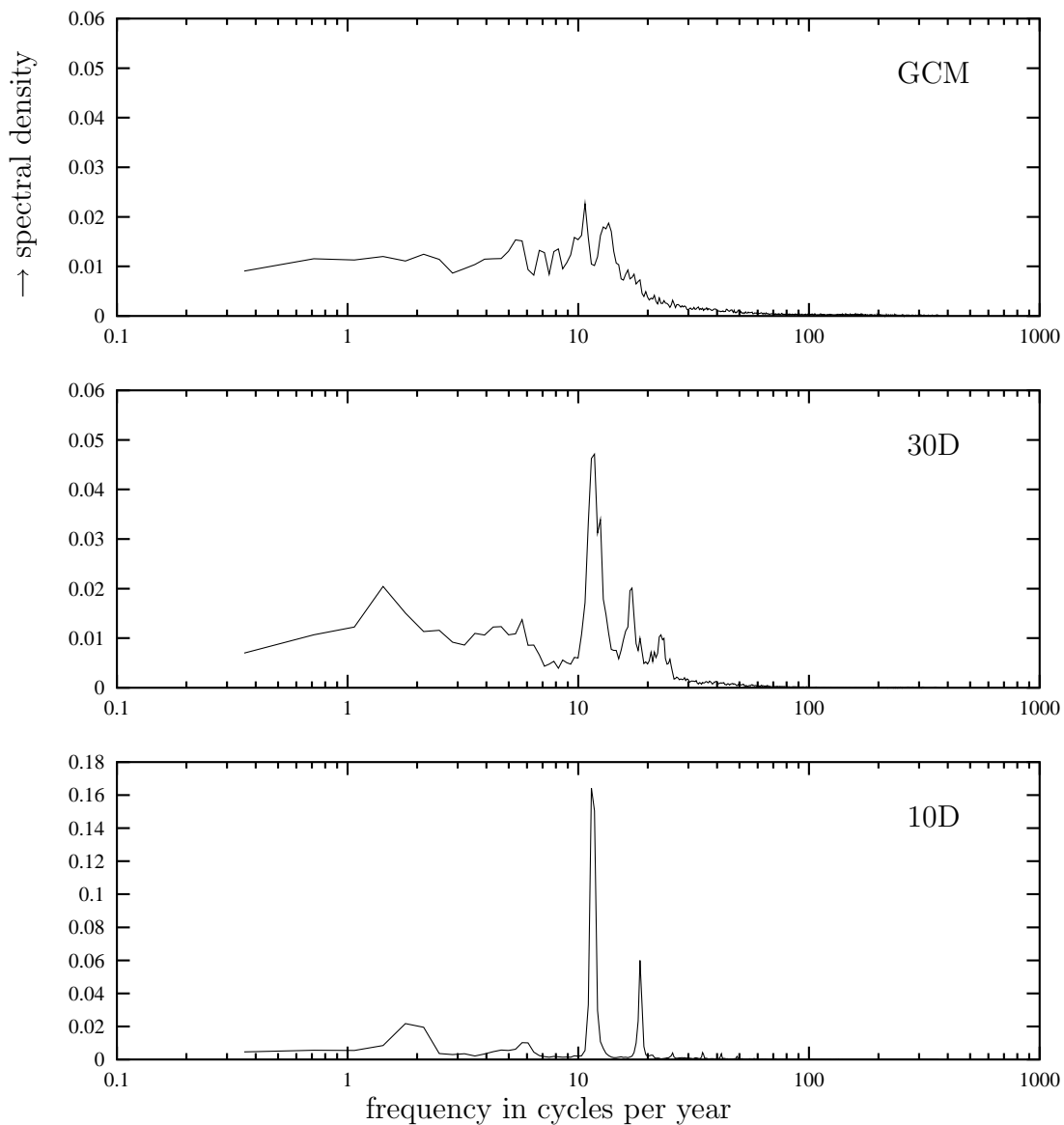

Figure 1.3: *Powerspectra of EOF1 in GCM, 30D and 10D models*

in both the 30D model and the GCM. Furthermore, its dimension is low enough to allow for the use of (some) techniques and concepts from dynamical system analysis.

1.4 Analysis of the 10D model

In the previous section, it was already noted that the 10D model behaviour is marked by outbursts of turbulent energy: phases with rather regular behaviour and low turbulent energy (so-called *laminar* phases) are followed by phases with chaotic behaviour and high turbulent energy, as can be seen in figure 1.1. The regularity of the behaviour at low energies indicates the presence of an unstable equilibrium solution of the model equations (1.4) close to $\mathbf{a} = 0$.

The equilibrium (or fixed point) can be found by continuation (starting at $\mathbf{a} = 0, \mathbf{F} = 0$), using e.g. the continuation software-package AUTO (Doedel and Wang (1995)), or using a rootfinder. With both methods the equilibrium point $\mathbf{a}_{eq}$ is easily found. Linearisation shows that it is a saddle-type fixed point, with 1 unstable and 4 stable complex pairs of eigenvalues. In table 1.1 the eigenvalues are listed, together with their associated periods $2\pi/|\text{Im}\,\lambda_i|$ and e-folding times $1/|\text{Re}\,\lambda_i|$. The periods of the five eigenmodes explain most of the peaks in the EOF1 powerspectrum (figure 1.3): those at frequencies of 2.0, 5.8, 11.5 and 18.6 cycles per year match very well with the periods of modes 5, 4, 3 and 2, respectively. The period of mode 1 is not visible in the EOF1 powerspectrum but can be observed in the powerspectra of most other EOFs (not shown).

The atmospheric patterns associated to the eigenmodes are shown in figure 1.4. Mode 1 is a northern hemispheric wavenumber 5 zonal wave at mid-latitudes. Modes 2 and 3 are equatorial waves; they are manifestations of the models' representation of the Madden-Julian Oscillation (MJO), also known as the 30-60 day tropical oscillation. Mode 2 has wavenumber 2 (an aspect of the MJO that has been observed before (Lau and Peng (1987), Salby and Hendon (1994)) but that is less well-known than the wavenumber 1 pattern of the MJO); it also shows some north-pacific activity. Mode 3 is a wavenumber 1 equatorial wave, the most famous aspect of the MJO (see Von Storch et. al. (1988)). Modes 4 and 5 are both largely northern-hemispheric; they resemble the Pacific/North-American (PNA) pattern. Mode 5 also shows an alternation between blocked and zonal flow in the north-atlantic region.

The e-folding times of the 5 eigenmodes may seem somewhat strange. They are one to two orders of magnitude larger than the most commonly known timescales associated

no.	eigenvalue		period	e-folding time
1	-0.00272154	$\pm\,0.438839\,i$	14.3 days	367.4 days
2	-0.00223039	$\pm\,0.319469\,i$	19.7	448.4
3	-0.00168416	$\pm\,0.198707\,i$	31.6	593.8
4	-0.00286361	$\pm\,0.0976673\,i$	64.3	349.2
5	0.00165548	$\pm\,0.0353438\,i$	177.8	604.1

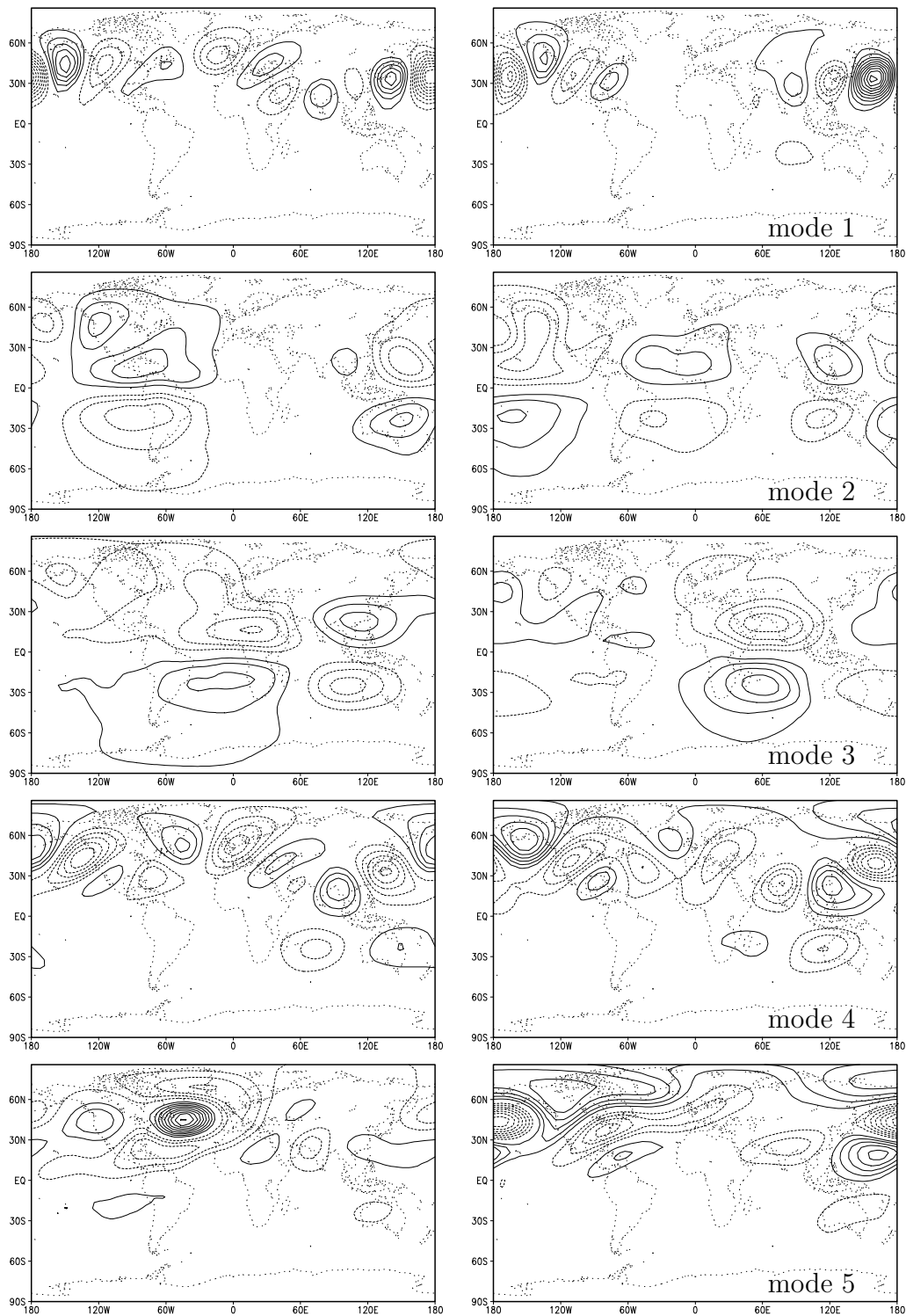Table 1.1: *Eigenvalues of linearised 10D EOF model*

Figure 1.4: *Eigenmodes of the 10D EOF-model linearised around the equilibrium state. Shown are eigenmode patterns (200 mb) at the start (left) and at a quarter (right) of one complete cycle. Contour intervals are 0.05 in non-dimensional units.*

to the growth and decay of atmospheric perturbations (typically a few days to a few weeks). This is due to the empirical nature of the model under consideration. Growth and decay cannot be attributed to physical processes (such as baroclinic instability or Ekman dissipation) anymore; rather, they combine the cumulative effect of many such processes. The empirical closure scheme applied by AB results in a model that describes the net effect of many different physical processes on the resolved patterns.

Another peculiarity of the model stems from the fact that it describes deviations from the mean flow instead of total atmospheric flow. By shifting the variables, say $\boldsymbol{\psi}$, to $\boldsymbol{\psi}' = \boldsymbol{\psi} - \overline{\boldsymbol{\psi}}$ (with $\overline{\boldsymbol{\psi}}$ the mean state), the quadratic nonlinearities (advection terms) $\mathsf{N}(\boldsymbol{\psi}, \boldsymbol{\psi})$ split into a nonlinear part $\mathsf{N}(\boldsymbol{\psi}', \boldsymbol{\psi}')$, a constant part $\mathsf{N}(\overline{\boldsymbol{\psi}}, \overline{\boldsymbol{\psi}})$ and a linear part $2\,\mathsf{N}(\overline{\boldsymbol{\psi}}, \boldsymbol{\psi}')$. The latter is added to the 'old' linear operator, which contained e.g. Coriolis terms, friction and dissipation. The new linear operator can have growing eigenmodes, whereas the old operator was purely dissipative. The energy transfer between the mean flow and its disturbances, which is modelled with nonlinear terms in total flow models, turns into a *linear* effect in the model of anomalous flow. As said, the linear operator of the anomaly-model can be unstable, i.e. can have eigenvalues with positive real part. This is the case for the 10D EOF model that is used here.

We rewrite the model equations (1.4) in terms of the eigenmodes of $\mathbf{a}_{\mathrm{eq}}$, yielding a system of 5 growing/decaying oscillations that are only nonlinearly coupled. Substituting $\mathbf{a} = \mathbf{a}_{\mathrm{eq}} + \mathsf{S}\,\mathbf{x}$ in (1.4) results in

$$\dot{\mathbf{x}} = \mathsf{B}\,\mathbf{x} + \tilde{\mathsf{N}}(\mathbf{x}, \mathbf{x}). \tag{1.5}$$

The (non-orthogonal) matrix S was chosen such that the new linear operator B is block-diagonal: its only nonzero elements are five blocks $\begin{pmatrix} \rho_i & -\omega_i \\ \omega_i & \rho_i \end{pmatrix}$ on the diagonal, with ρ_i and ω_i the real and imaginary parts of the ith eigenvalue pair. The variables $\mathbf{x}$ and the parameters of the equation are all real; the jth eigenmode is spanned by x_{2j-1} and x_{2j}. Turbulent energy in terms of $\mathbf{x}$ reads $E = (\mathbf{a}_{\mathrm{eq}} + \mathsf{S}\,\mathbf{x})^{\mathrm{T}}\,(\mathbf{a}_{\mathrm{eq}} + \mathsf{S}\,\mathbf{x}) \neq \mathbf{x}^{\mathrm{T}}\,\mathbf{x} + \mathbf{a}_{\mathrm{eq}}^{\mathrm{T}}\,\mathbf{a}_{\mathrm{eq}}$.

Numerical integration of the model reveals that the linear terms $\mathsf{B}\,\mathbf{x}$ dominate the behaviour of the model. The full model $\dot{\mathbf{x}} = \mathsf{B}\,\mathbf{x} + \tilde{\mathsf{N}}(\mathbf{x}, \mathbf{x})$ can be seen as a simple, linear system $\dot{\mathbf{x}} = \mathsf{B}\,\mathbf{x}$ perturbed by relatively small nonlinear terms $\tilde{\mathsf{N}}(\mathbf{x}, \mathbf{x})$. Averaged in time, the quantity $|\tilde{\mathsf{N}}(\mathbf{x}, \mathbf{x})|/|\mathsf{B}\,\mathbf{x}|$ is about 1/4. The unperturbed (linear) system already explains much about the intrinsic frequencies of the full (i.e. perturbed) system. However, it does not explain the spectral peak at a frequency of about 0.15 cycles per year. This peak is apparently due to the nonlinear nature of the full model. We therefore need to consider the full nonlinear model in order to understand the ultra-low frequencies that are generated.

A clue as to what is determining the global dynamics of the full model is provided by figure 1.5, in which the squared norm $|\mathbf{x}|^2$ is plotted against time during an integration of the full 10D model. This quantity is comparable, but not equal, to the turbulent energy. It can be seen that the system approaches the point $\mathbf{x} = 0$ (i.e. the equilibrium $\mathbf{a}_{\mathrm{eq}}$ when using $\mathbf{a}$ as a variable) very closely, time and again. Such behaviour suggests the existence of a homoclinic orbit: a phase space trajectory which connects an (unstable) equilibrium point with itself. The presence of such an orbit would provide a good explanation for the departure from and return to (the vicinity of) the equilibrium point by the system, as is
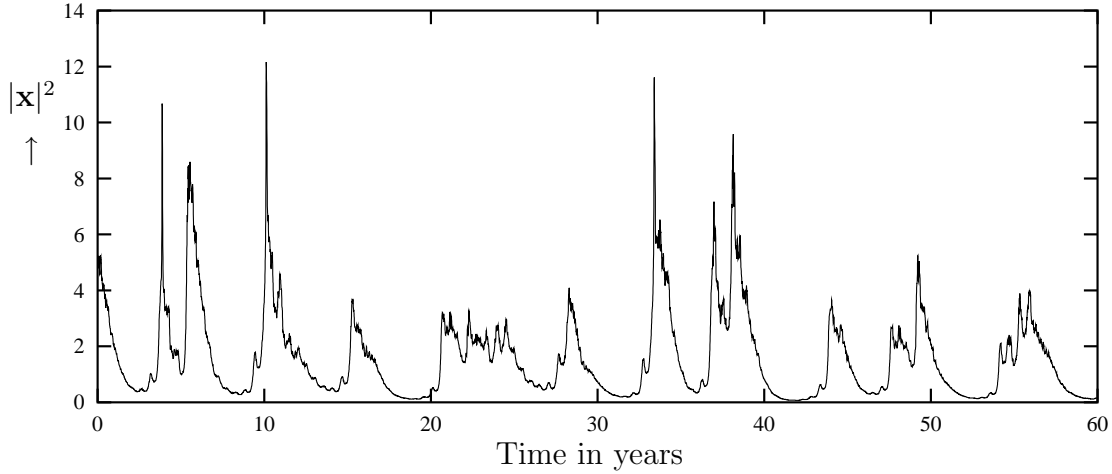
Figure 1.5: *Timeseries of squared norm ($|\mathbf{x}|^2$) of phase space distance to equilibrium point*

visible in the 10D model, but also, less pronounced, in the 30D EOF model and in the GCM.

We shall investigate the hypothesis that the global, long timescale behaviour of the 10D model is governed by the presence of a homoclinic orbit in the following section, after providing some theory on homoclinic orbits. We do not attempt to detect this homoclinic orbit in the 10D model (this is a very hard task in 10 dimensions), but rather we focus on a 4D subsystem of the 10D system which contains the essence of the long timescale dynamics of the 10D model.

1.5 A 4D chaos generator

Given a set of ODEs $\dot{\mathbf{y}} = \mathbf{G}(\mathbf{y}, \boldsymbol{\mu})$ with $\boldsymbol{\mu}$ denoting the vector of parameters, a *homoclinic orbit* $\mathbf{y}_{ho}(t)$ is the projection onto phase space of a non-trivial solution of the ODEs which asymptotically approaches a fixed point $\mathbf{y}_0$:

$$\lim_{t \to +\infty} \mathbf{y}_{ho}(t) = \lim_{t \to -\infty} \mathbf{y}_{ho}(t) = \mathbf{y}_0 \tag{1.6}$$

With $\mathbf{y}_{ho}(t)$ being non-trivial we mean that $\mathbf{y}_{ho}(t) \neq \mathbf{y}_0$ for some values of t. A *heteroclinic orbit* (or connection) $\mathbf{y}_{he}(t)$ approaches in forward and backward time two different fixed points $\mathbf{y}_+$ and $\mathbf{y}_-(\neq \mathbf{y}_+)$:

$$\lim_{t \to +\infty} \mathbf{y}_{he}(t) = \mathbf{y}_+, \quad \lim_{t \to -\infty} \mathbf{y}_{he}(t) = \mathbf{y}_- \tag{1.7}$$

A *heteroclinic cycle* is a set of heteroclinic orbits forming a closed loop. The simplest version consists of two heteroclinic orbits: one running from one fixed point $\mathbf{y}_-$ to another $\mathbf{y}_+$, the second running from $\mathbf{y}_+$ back to $\mathbf{y}_-$. Strictly speaking, homoclinic and heteroclinic orbits can also exist between other solutions than fixed points (e.g. periodic solutions, invariant tori etc.), but we will not treat those possibilities here.

Let us focus on the homoclinic case and assume that $\mathbf{y}_{ho}(t)$ exists for the value $\boldsymbol{\mu} = \boldsymbol{\mu}_{ho}$ of the parameter(s). Generically (assuming nondegenerate and nonzero eigenvalues), there

are three possibilities for the leading eigenvalues of the system linearised around $\mathbf{y}_0$:

1. Two reals: (λ_1, λ_2) with $\lambda_1 < 0 < \lambda_2$ (the saddle case)
2. One real, one complex pair: $(\lambda_1, \lambda_2 \pm i\omega_2)$ with either $\lambda_1 < 0 < \lambda_2$ or $\lambda_2 < 0 < \lambda_1$ (the saddle-focus case)
3. Two complex pairs: $(\lambda_1 \pm i\omega_1, \lambda_2 \pm i\omega_2)$ with $\lambda_1 < 0 < \lambda_2$ (the double-focus or bifocal case)

In our 10D model, the eigenvalue spectrum of the fixed point consists exclusively of complex pairs, so the eventual homoclinic orbit connected to this fixed point should be of the bifocal type. For the bifocal case it is known that at the bifurcation point $\boldsymbol{\mu}_{ho}$, as well as in a parameter neighbourhood of $\boldsymbol{\mu}_{ho}$, a countable infinity of periodic orbits (up to arbitrarily high periods) and an uncountable number of aperiodic orbits exist. All these orbits are of saddle-type (i.e. unstable in some phase-space directions). A horseshoe-mapping can be constructed, which means that a chaotic invariant set (consisting of the collection of unstable (a)periodic orbits) is created. In our case, the unstable eigenvalues lie closer to the imaginary axis than the stable eigenvalues and most evidence from dynamical systems suggests that the chaotic invariant set is then attracting. In other words, from the eigenvalue-spectrum and the (conjectured) existence of the homoclinic orbit, it can be inferred that a chaotic attractor exists for the model under consideration. More detailed information on bifocal homoclinic orbits can be found in Fowler and Sparrow (1991), Glendinning (1997) and Laing and Glendinning (1997). For a more general treatment of homoclinic and heteroclinic bifurcations, see e.g. Kuznetsov (1995) or Wiggins (1988).

It must be stressed that a homoclinic bifurcation is a codimension 1 phenomenon, which means that in general one parameter of the system must be tuned in order to arrive at the bifurcation point. The homoclinic orbit exists exactly at this bifurcation point. Away from this point, the homoclinic orbit disappears, but its influence can still be felt by the system. Relating the behaviour of a system to homoclinic dynamics must therefore be understood as identifying a homoclinic bifurcation point at parameter values close enough to the values one is interested in, so that the influence of the homoclinic orbit is still present.

Returning to our 10D atmosphere model, it can be seen that the interpretation of its long timescale behaviour in terms of homoclinic dynamics is rather natural: the model possesses a saddle fixed point and returns to this point time and again. In more physical terms, one could characterise the long timescale behaviour as follows: the system comes close to the fixed point $\mathbf{x} = 0$; the unstable eigenmode starts growing. Once it has acquired substantial energy, this energy is transferred, via nonlinear interactions, to other modes. Eventually these other modes dissipate their energy until the system is back in the vicinity of the fixed point. Then the whole cycle more or less repeats itself. Both the growth and the decay take place in an oscillating manner.

We wil try to represent the crucial aspects of this scenario (2-dimensional oscillating growth – transfer of energy – 8-dimensional oscillating decay) by a lower-dimensional system. The minimum that is needed for such behaviour must be 4 dimensions: 2 to produce a growing oscillation, 2 for a decaying oscillation. Such a system can be derived from the 10D system in a crude way by replacing the 4 dissipating modes by just one: we

can isolate a 4D subsystem from the 10D model, consisting of mode 5 (the only growing mode) and one of modes 1-4, and their linear and nonlinear (self-) interactions. The effect of the dissipating mode has then to be identified with the cumulative effect of modes 1-4 in the 10D model; mode 5 is of course identified with itself. It is not necessary to restrict ourselves to triad interactions (as in spectral models) since each interaction between two EOFs already involves many triad interactions between spectral modes.

The choice of the dissipating mode is nontrivial. From the point of view of dynamical systems theory, selecting the leading eigenmodes, hence modes 3 and 5, would be the most natural choice, since the leading eigenmodes determine the geometry of the (phase space) flow close to the fixed point. However, the system consisting of modes 3 and 5 and their interactions doesn't yield interesting behaviour - numerical integration reveals that it ends up in a nontrivial stable fixed point. The combinations 1-5 and 2-5 don't work either: the former also goes to a nontrivial fixed point, the latter goes to unphysically high turbulent energy (a factor 10^5 too high). The only combination that gives similar results as the 10D model is that of modes 4 and 5. The 4D model of these two modes and their interactions reproduces the cycle of growth-transfer-decay; it returns to the fixed point in the origin of phase space, again and again, just like the 10D model. Figure 1.6 makes clear how the total energy is distributed over the dissipating and the growing modes during a burst, both for the 10D model and the 4D model.

Several reasons can be given why the selection of mode 4 gives good results and the selection of the leading stable eigenmode (mode 3) doesn't. A physical argument is that modes 4 and 5 are both mainly northern-hemispheric modes, in contrast, mode 3 is mainly tropical. This makes the interaction between 4 and 5 easier than between 3 and 5. From a more mathematical point of view one can observe that the frequencies of modes 4 and 5 are rather close to a 1:3 resonance (i.e. strong resonance) whereas the frequencies of modes 3 and 5 tend towards 1:6, i.e. weak, resonance. Strong resonances can be essential for the dynamics of a system. From the theory of nonlinear resonant interactions it is known that, in general, couplings of modes at (or near) low-order (i.e. strong) resonance are the most efficient interactions (see e.g. Arnol'd (1983)). A precise clarification of the influence of the 1:3 resonance is outside the scope of this paper; given the small real parts of the eigenvalues, such a clarification would require knowledge, currently unavailable, of the unfolding of a double Hopf-bifurcation at 1:3 resonance (see e.g. LeBlanc (2000) and Kuznetsov (1995)).

An inspection of the amplitudes of all 5 modes from an integration of the 10D model (not shown) makes clear that mode 4 plays an important role: in the onset of each burst it is the first of the stable modes to be excited by the unstable mode 5. Apparently, the latter transfers its energy more easily to mode 4 than to modes 1,2,3. In the decaying phase of each burst, the decay timescale cannot be clearly associated with the stable eigenvalues: the decay is much faster than one would expect from the real parts of the eigenvalues. During the bursts the system is probably too far away from the fixed point for the linear regime to be valid. Further experimentation confirms that mode 4 is crucial in the maintenance of the energy cycle (growth-transfer-decay): the 8D system of modes 1,2,3,5 does not reproduce the cycle (integration results not shown), the 4D system of modes 4 and 5 does.

The 4D subsystem is given by

$$
\begin{aligned}
\dot{x}_7 &= \rho_4 x_7 - \omega_4 x_8 + N_7(x_7, x_8, x_9, x_{10}) \\
\dot{x}_8 &= \rho_4 x_8 + \omega_4 x_7 + N_8(x_7, x_8, x_9, x_{10}) \\
\dot{x}_9 &= \rho_5 x_9 - \omega_5 x_{10} + N_9(x_7, x_8, x_9, x_{10}) \\
\dot{x}_{10} &= \rho_5 x_{10} + \omega_5 x_9 + N_{10}(x_7, x_8, x_9, x_{10})
\end{aligned}
\tag{1.8}
$$

x_7, x_8 correspond to mode 4 and x_9, x_{10} to mode 5. The nonlinearities N_i consist again only of quadratic terms. All parameters of the above equations are numerically fixed, due to the empirical nature of the EOF-model. If we want to perform a bifurcation analysis we will have to pick our own bifurcation parameter, since common parameters like the
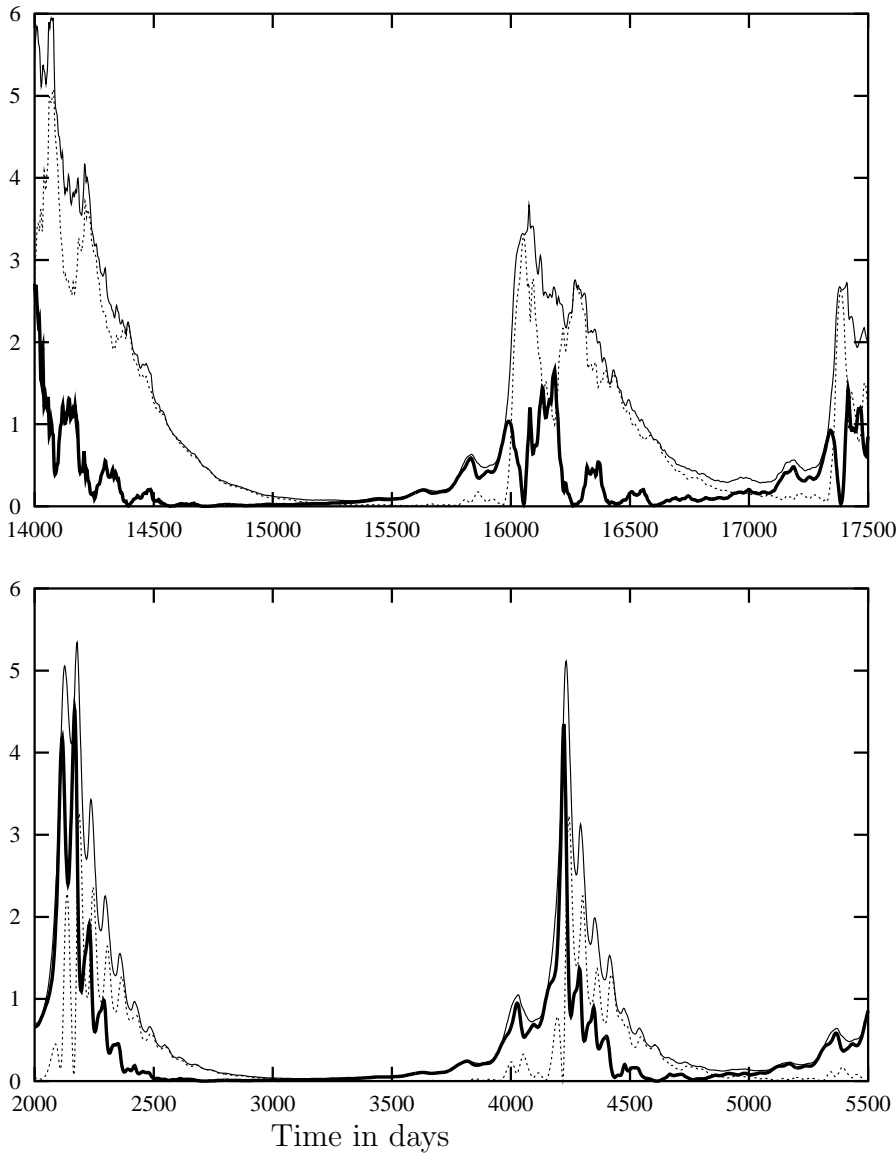

Figure 1.6: *Typical cycle of growth-transfer-decay in 10D model (top) and 4D model (bottom). Plotted are the squared norms of all modes together (normal line), of the growing mode (thick line) and of the dissipating modes together (dotted line)*

Ekman damping rate or orographic scale height are not available anymore. We make the choice of altering ω_5, which has the effect of altering the frequency resonance relation between the two modes. There is no clear physical effect associated with such a change (because of the empirical nature of the model, any chosen parameter will be difficult to interpret physically); however, a changing frequency-ratio is likely to cause changes in the model behaviour, thereby providing a better understanding of that behaviour (it must be mentioned that a frequency-altering parameter also appears as one of the three unfolding parameters for the strongly resonant double Hopf-bifurcation in LeBlanc (2000)).

Phase portraits (projected onto the (x_9, x_{10}) plane) of numerical integrations with varying ω_5 are shown in figure 1.7 (leaving out initial transient motion). The upper left panel shows the behaviour at $\omega_5 = 0.0353438$, the original value coming from the EOF model. In figure 1.8 timeseries of the squared L_2-norm, $|(x_7, x_8, x_9, x_{10})|^2$, are shown at the same parameter values. Both chaotic and rather regular (even completely periodic) behaviour is found. To get a more detailed picture of the behaviour at various values of ω_5, Poincaré-sections were made in the halfplane $(x_9 < 0, x_{10} = 0)$. A plot of the x_9-value of the intersection points with varying ω_5 (a so-called limit-point diagram, see fig. 1.9) exhibits the existence of both chaotic and periodic windows.

The most interesting region of parameter space is the region of decreasing ω_5, i.e. $\omega_5 < 0.0353438$. The model behaviour starts to show more and more pronounced bursts, with long periods in between, during which the systems resides very close to the fixed point in the origin (the squared phase space norm dropping below 0.005, see figure 1.8). The behaviour stays chaotic but nevertheless becomes very structured. The recurrence of the system, after each burst, to the fixed point in the origin, and the very long time spent extremely close to this fixed point, provides strong evidence for our hypothesis that a homoclinic orbit underlies the observed model behaviour.

By carefully inspecting figure 1.9 it was possible to identify an attracting periodic orbit, at $\omega_5 = 0.0290188$. This orbit, phase-space projections of which are shown in figure 1.10, has a period of 3394 days; it provides a good approximation to the homoclinic orbit. A 1-parameter continuation of the periodic orbit, using AUTO, in order to find the precise location of the homoclinic bifurcation (the point in parameter space where the period of the periodic solution approaches infinity) failed to be conclusive, probably because of numerical problems due to the complex shape of the periodic solution.

The shape of the periodic orbit of figure 1.10 hints at the existence of a second fixed point. This point indeed exists; it has coordinates $(x_7, x_8, x_9, x_{10}) = (0.413562, -0.650201, 1.64348, 1.09727)$ when $\omega_5 = 0.0290188$ and is unstable. Its existence and location raise the question whether it is a heteroclinic cycle instead of a homoclinic orbit that structures the system behaviour. There are two reasons to hold on to the hypothesis of homoclinic dynamics. First, the time spent near the fixed point in the origin is many times longer (and the proximity many times closer) than the time spent near the second fixed point. At the point $\omega_5 = 0.0353438$ in parameter space, the fixed point in the origin clearly still plays an important role for the behaviour of the system, whereas there seems to be no trace of the other fixed point in the model behaviour. Second, since a heteroclinic cycle involves two connections between fixed points, two parameters must in general be tuned in order to arrive at a heteroclinic cycle, rather than one, as is the case for a homoclinic orbit (a heteroclinic cycle is a codimension 2 phenomenon, a homoclinic orbit codimension 1). It
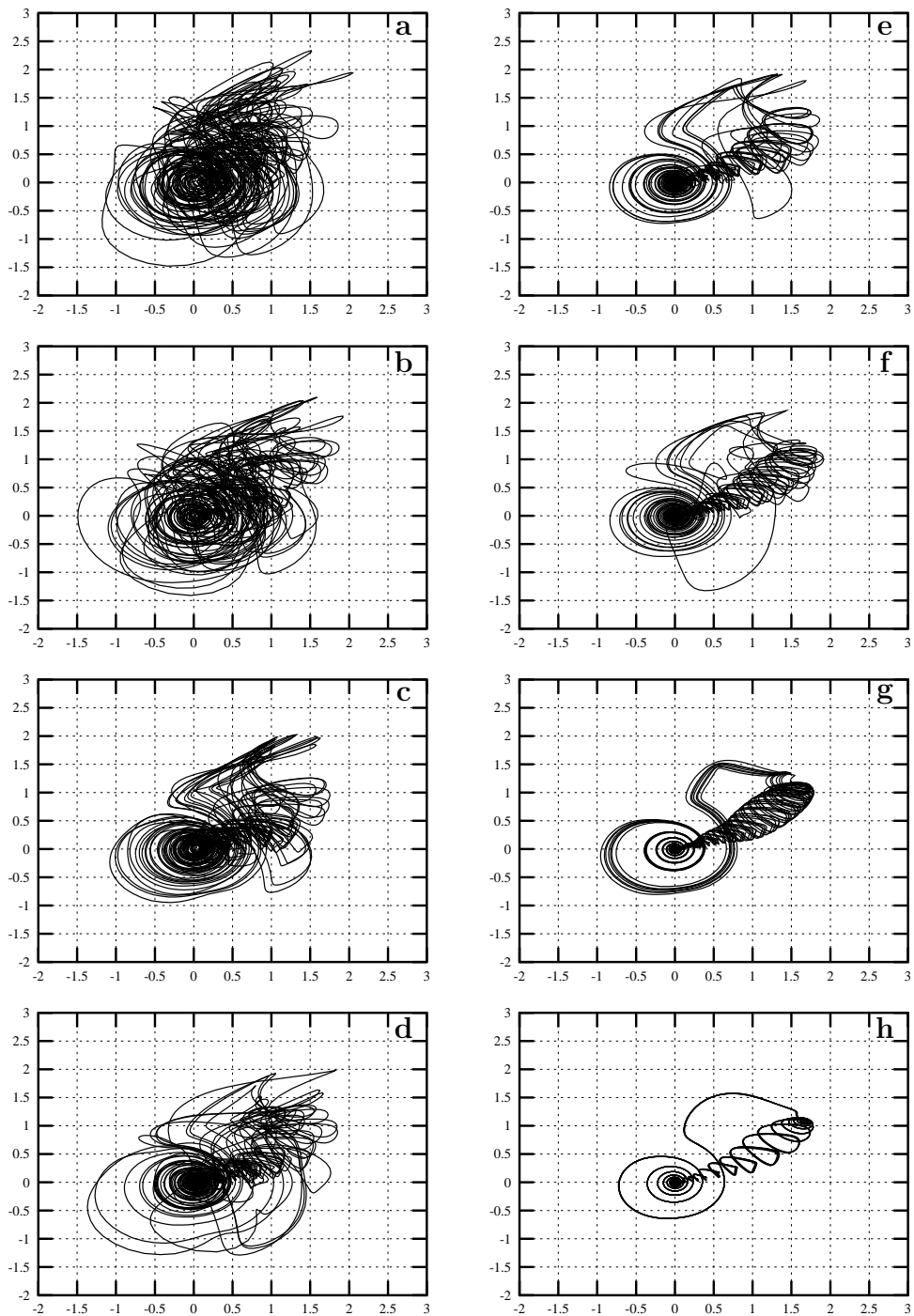
Figure 1.7: *Phase plots (x_9 versus x_{10}) of 4D model at different values of bifurcation parameter ω_5: 0.0353438 (a), 0.0343438 (b), 0.0333438 (c), 0.0323438 (d), 0.0313438 (e), 0.0303438 (f), 0.0293438 (g), 0.0290188 (h). Integration time is 20 000 days.*
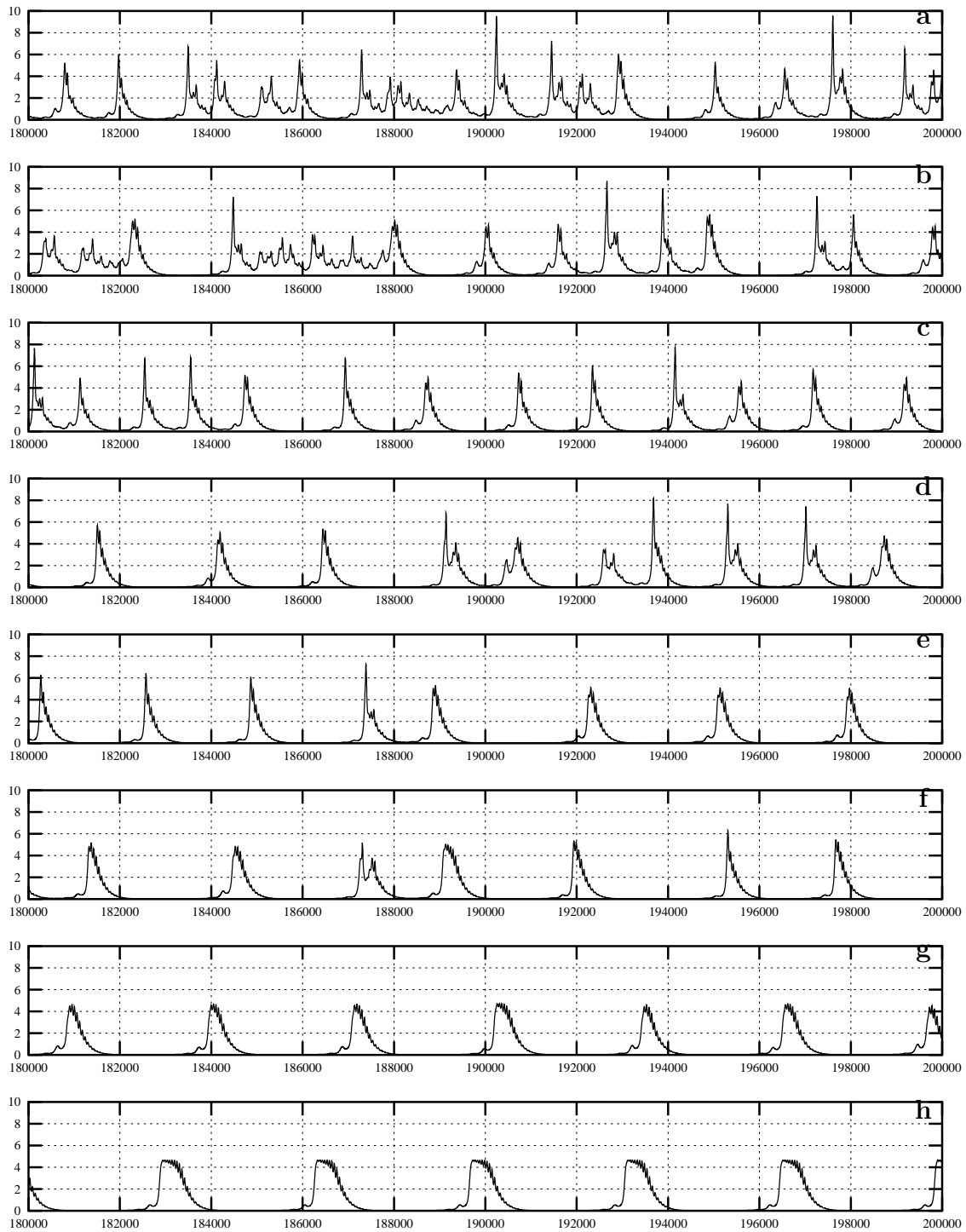
Figure 1.8: *Timeseries of squared phase space norm ($|(x_7, x_8, x_9, x_{10})|^2$) of 4D model at different values of bifurcation parameter ω_5 (see figure 1.7 for numerical values). Integration time is 20 000 days.*

is therefore improbable to have encountered a heteroclinic cycle, since only one parameter was varied. It is possible for a heteroclinic cycle to become a codimension 1 phenomenon in a system which possesses symmetry, but our 4D model doesn't appear to have any symmetry. Nevertheless, it is possible that our model is close to some symmetry that decreases the codimension of a heteroclinic cycle. Breaking such a symmetry can very well break the heteroclinic cycle and produce a homoclinic orbit instead. This interesting possibility lies, however, beyond the scope of this paper.

In conclusion, the existence of a homoclinic orbit of the bifocal type in the 4D model explains the chaotic behaviour and the occurrence of very long timescales in the system. The structure of the orbit also explains the observed cycle of growth, transfer and decay in the system: it is the cycle of the homoclinic orbit which starts at the origin, moves away and eventually returns to the equilibrium point. The attractor generated around the homoclinic solution will possess roughly this same structure, so the behaviour of the system, as it evolves on the attractor, will be marked by the growth-transfer-decay cycle. This is in agreement with our earlier observations (e.g. figure 1.6).

1.6 Homoclinicity in complex models

A natural question to ask is whether the ULFV-producing mechanism found for the low-order model (generation of chaos due to a bifocal homoclinic orbit attached to an
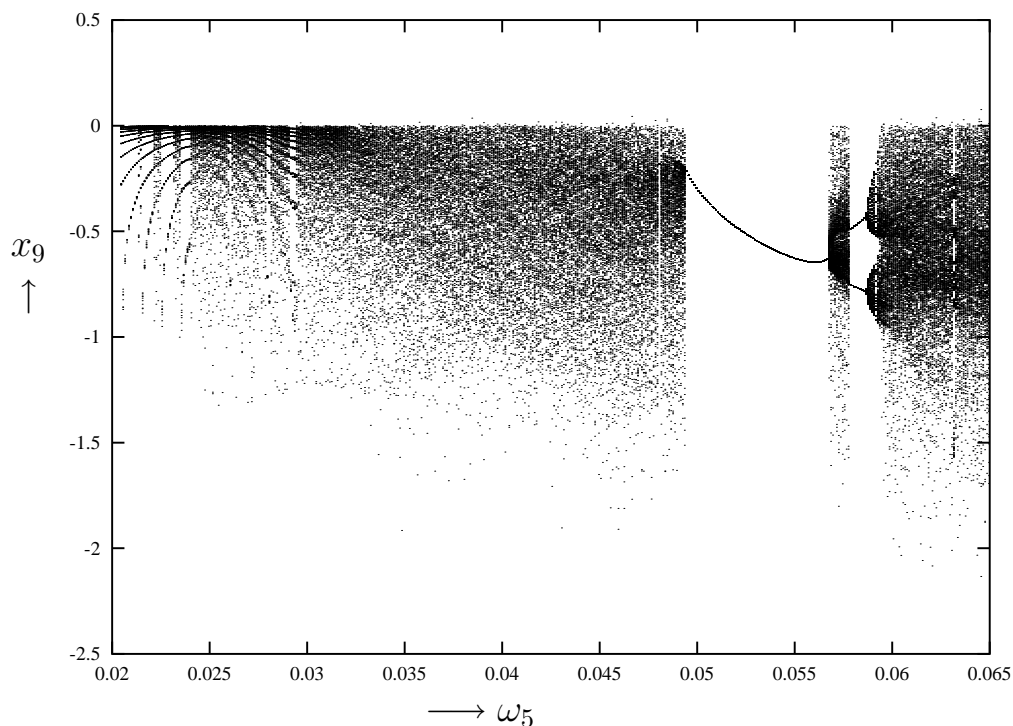
Figure 1.9: *Limit-point diagram showing the result of Poincaré sections made at various values of the bifurcation parameter ω_5.*

equilibrium state close to the climate mean state) will play a role in more complex models. Will the 30-dimensional version of the empirical EOF-model and the GCM possess such an orbit? Obviously, establishing direct numerical evidence for such a conjecture is hardly possible. Nevertheless, the situation for the 4D model is quite generic. The model describes deviations from an equilibrium state close to the atmospheric mean state. The equilibrium is of saddle type, it has both growing and decaying eigenmodes. This mixed
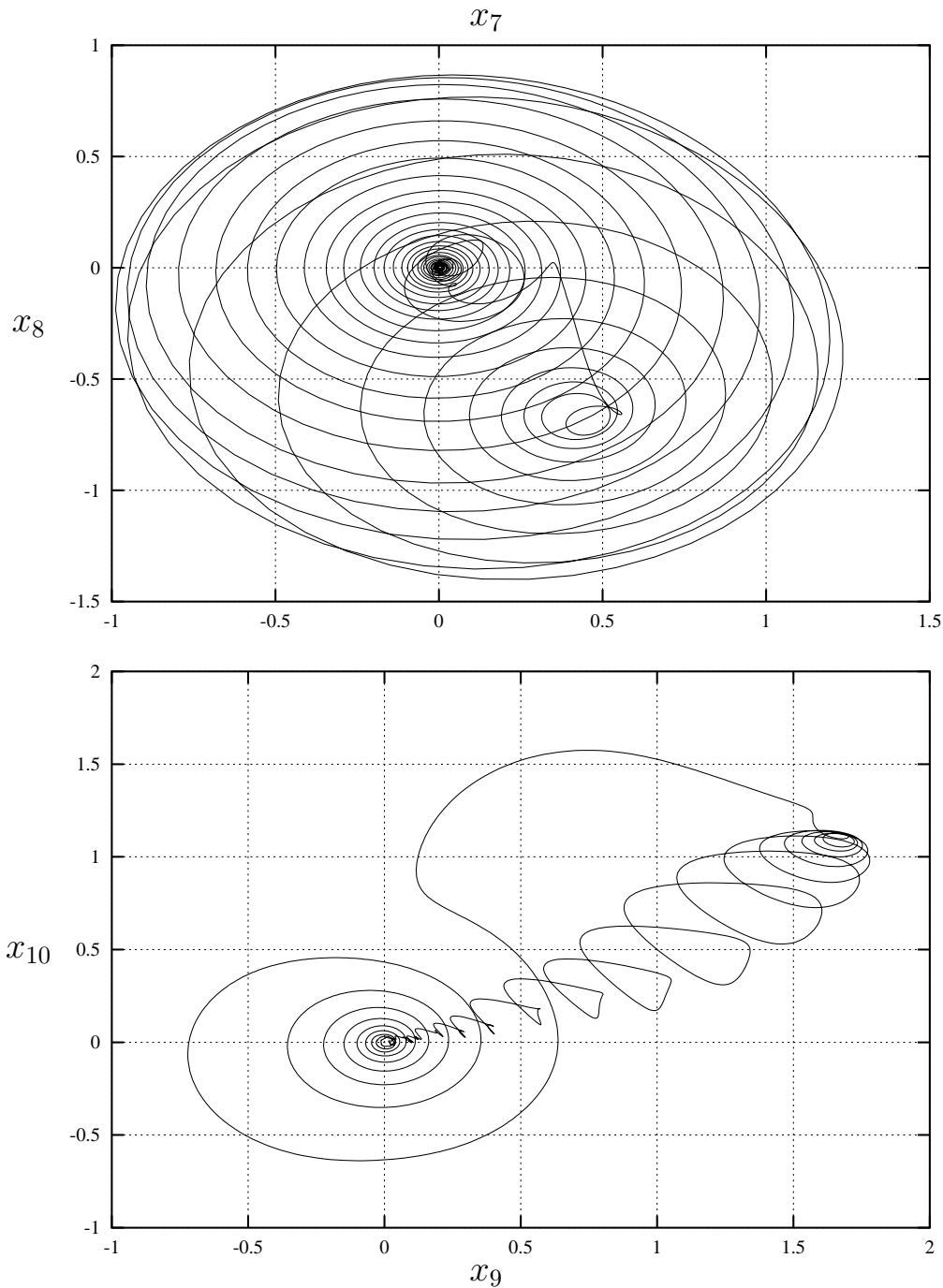
Figure 1.10: *Phase space projections of the attracting periodic orbit at $\omega_5 = 0.0290188$. Period = 3394 days*

stability does not depend on the complexity of the model. Moreover, an atmospheric system linearised around the equilibrium point will generically have complex pairs of eigenvalues, which will cause an eventual homoclinic orbit connected to this equilibrium to be of the bifocal type. Unlike the other two cases (the saddle case and the saddle-focus case), where certain conditions on the eigenvalue-spectrum have to be met in order to create chaotic dynamics, the bifocal case of homoclinic orbits always involves the generation of a chaotic invariant set. If this set is attracting, the system will possess chaotic dynamics and nontrivial behaviour on timescales of arbitrary length.

The existence of a homoclinic orbit for a model describing anomalous atmospheric flow remains, obviously, the *pièce de résistance* of this analysis. Proving, even numerically, the existence of a homoclinic orbit is a very hard task even in three- or four-dimensional models; for high-dimensional models such as GCMs it is virtually impossible. But carefully looking at the outcome of such complex models may provide some circumstantial evidence. Timeseries of the turbulent energy of the GCM in the first 30 EOFs (figure 1.1) clearly show significant changes in the distance of the system to the origin of phase space, i.e. to the mean state. The recurrence of the system to the vicinity of this mean state (which is likely to be close to some equilibrium solution) hints at the presence of a homoclinic connection. The simple fact that the system moves away from the fixed point due to growing perturbations on the mean flow, but eventually returns to the vicinity of the fixed point, already suggests homoclinicity – it is the only known dynamical mechanism of moving away and returning back to some equilibrium state (disregarding here, for brevity's sake, the possibility of a heteroclinic cycle). We therefore expect the creation of chaotic behaviour in the atmosphere due to homoclinicity to be a general mechanism, not an artefact of the model used.

However, we do not want to state here that homoclinic orbits are likely to be existent in truly high-dimensional, complex, realistic atmosphere models, which contain considerable small-scale detail. Rather, we hypothesise that it is the large-scale atmospheric circulation which is marked by homoclinic behaviour. The smaller-scale processes in the atmosphere are likely to destroy the supposed homoclinic connection, but the large-scale behaviour will still show traces of the homoclinic dynamics. This interpretation is supported by the comparison of the 10D and 30D EOF models and the GCM. The 10D model shows most clearly the characteristics of homoclinic dynamics; adding smaller-scale patterns obscures this behaviour, as can be seen in the 30D model. The GCM still shows large variations in turbulent energy and recurrence to the vicinity of the time mean state if its data is projected onto only the first 30 EOFs (i.e. the largest spatial scales). This recurrence is less clear if all resolved scales are taken into account. In short: the small spatial scales provide perturbations, "noise", which affect (disturb) the homoclinic behaviour of the large-scale atmospheric circulation.

Such interpretation is supported by the results of two previous studies, in which signs of homoclinic behaviour can be identified. James et. al. (1994) run their T21L5 model for 100 years and calculate EOFs of the zonal-mean zonal wind. Then, using an integration with imposed seven-fold symmetry in meridional direction (only wavenumbers 0, 7 and 14 are retained), phase-space projections onto the plane of the first two Principal Components are made which clearly show that the model returns time and again to the same neighbourhood (small positive PC1, PC2 almost zero). Note that this neighbourhood, at

least in that particular projection, is close to the time-mean state, just as in our results. Moreover, projection onto the first two EOFs implies a selection of large spatial scales. Such a selection, or filtering, is likely to enhance the visibility of the homoclinic behaviour.

In an entirely different study, directed at the existence of free modes, Branstator and Opsteegh (1989) find orbits recurring to one point in phase space when integrating the barotropic vorticity equation with forcing and dissipation, at T15 resolution. The point to which the system returns is a numerically approximated solution (i.e. equilibrium) of the free (unforced, nondissipative) barotropic vorticity equation. The system seems to be further influenced by another equilibrium state, but it doesn't approach this second fixed point as close as it approaches the first fixed point. This is reminiscent of our findings for the 4D model and raises again the question to what extent the supposed homoclinic behaviour is related to an eventual heteroclinic cycle between two equilibrium states.

It must be stressed that the above studies have a different focus than the one presented here. One major difference lies in the complexity of the models used. Therefore, our interpretation of certain results of James et. al. and Branstator and Opsteegh as signs of homoclinic behaviour does not necessarily imply that the same physical mechanism(s) can be associated with all three cases of supposed homoclinic dynamics. Rather, we want to draw attention to a mathematical structure which seems to be also present in other (and more complex) atmospheric models than the low-order EOF model of Achatz and Branstator. The physical properties associated with this mathematical structure need to be further elucidated.

1.7   A T21 barotropic model

To further support the hypothesis that homoclinic behaviour is a generic aspect of models of large-scale atmospheric flow, a standard barotropic spectral model is taken as an example. This model, truncated at T21, describes the winter circulation of the northern hemisphere and possesses a realistic climate and reasonable low-frequency variability. The model has 231 dimensions, due to its restriction to one hemisphere. Detailed information on the model can be found in Selten (1995). Here, we only mention the presence of Ekman damping, hyperviscosity (i.e. scale-selective damping) and orography in the model. The model is forced by a constant vorticity forcing which has been tuned in order to get a realistic climate mean state and realistic low-frequency variability.

The model has been integrated for 10 000 days, using a Runge-Kutta fourth-order routine with a 30 minutes timestep and daily output. The timescale of the Ekman damping is set to 15 days; the hyperviscosity is set such that the damping timescale for wavenumber 21 is 3 days. The time mean was calculated and subtracted from the output, resulting in timeseries containing deviations of spectral coefficients from their time mean value. Using these data, we followed essentially the same steps as previously for the EOF-model. A kinetic energy metric was used to calculate the turbulent energy of the model run (i.e. the phase space distance to the climate mean state was calculated using an energy norm). The timeseries of the turbulent energy is shown in figure 1.11. As before, large variations in turbulent energy can be seen to occur. The model returns, time and again, to the vicinity of the climate mean state.

Furthermore, phase plots were made from one segment of the data: between day 1550 and day 2000, clear transitions in turbulent energy can be observed (also shown in figure 1.11). A projection of the phase space orbit of the model onto the plane spanned by the real parts of the Y_3^2 and Y_5^2 spectral coefficients (picked, somewhat arbitrarily, to focus
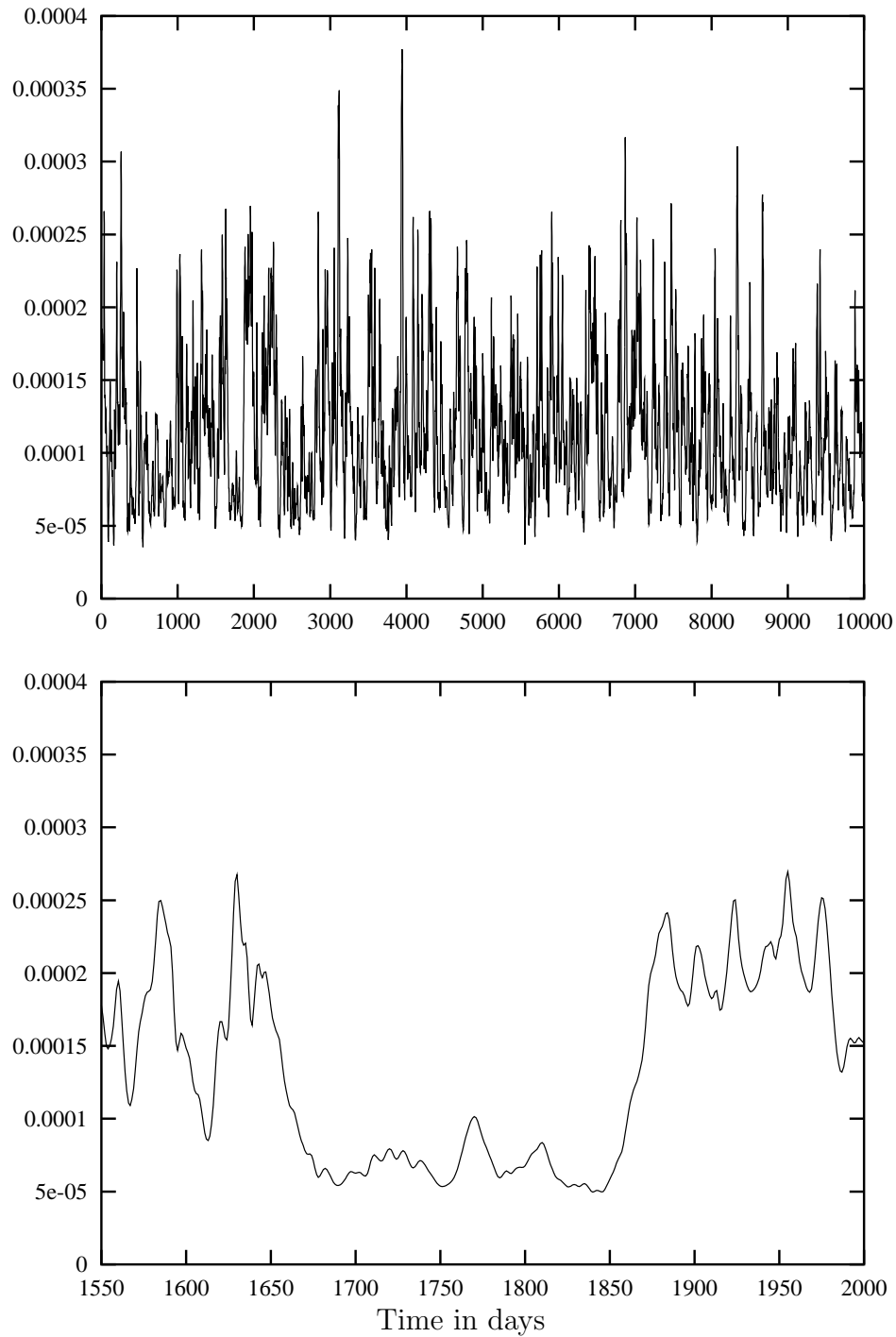

Figure 1.11: *Turbulent energy during 10000 days integration of T21 standard barotropic model*

on large-scale patterns) is shown in figure 1.12. To further clarify what is happening, the segment has been cut in three: days 1550-1680 (high turbulent energy), days 1680-1850 (low turbulent energy) and days 1850-2000 (again high turbulent kinetic energy). The projection of these three separate segments are also shown in figure 1.12. The plots clearly show a "trapping" of the system in a rather small region close to the origin (the climate mean state) during days 1680-1850, indicating the presence of a fixed point near which orbits remain for some time. Also, the other two segments (days 1550-1680 and 1850-2000) show large excursions in one specific direction of phase space. Apparently, the excursions are not completely arbitrary but somehow structured; a homoclinic orbit could explain such structure.

The overall picture suggests the influence of a homoclinic orbit in this T21 model of barotropic flow, supporting our hypothesis that the presence of homoclinic dynamics in models of large-scale atmospheric flow is a robust feature, not a model artefact. We
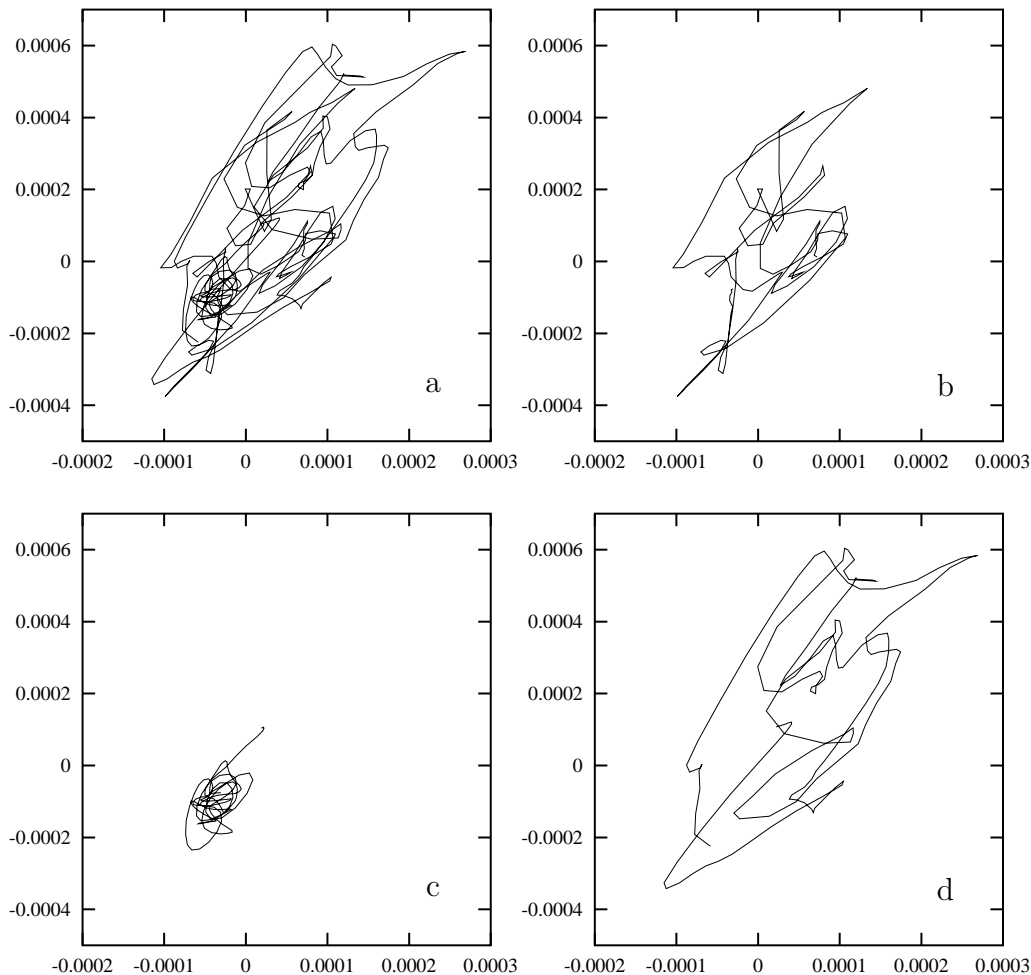

Figure 1.12: *Phase space projections of the T21 barotropic model onto the plane of $Re(Y_3^2)$ (horizontal) and $Re(Y_5^2)$ (vertical). a: days 1550-2000; b: days 1550-1680; c: days 1680-1850; d: days 1850-2000*

expect to report in more detail about the dynamics of the T21 model in the near future.

1.8   Conclusion

We have studied the dynamics of an efficient 10-dimensional EOF model describing global atmospheric behaviour, and found the dynamics on long timescales to be essentially driven by the presence of a homoclinic orbit connecting the equilibrium state close to the climate mean state with itself. The orbit lies at the core of the growth-transfer-decay cycle observed in the model. This cycle is overpronounced in the model used (due to its simplicity) but is likely to be also present in more complex models of atmospheric flow. Circumstantial evidence for this conjecture can be found in some previous studies, as well as in the observed variances in atmospheric turbulent energy in the 30-dimensional version of the EOF model and in the GCM the EOF model is based on. Further support for these findings comes from an integration of a T21 standard (forced, dissipative) barotropic model for northern hemisphere atmospheric flow. Both phase space projections and timeseries of turbulent energy from this model show signs of homoclinic behaviour.

The existence of a homoclinic orbit of the described type (bifocal) always entrains the creation of chaotic behaviour, and could thereby explain the ultra-low frequency variablity (ULFV) present in "stand-alone" atmospheric models (without coupling to oceans or other slow climate components).

The analysis and results obtained in this study raise some interesting issues that warrant further study. More evidence for the presence of homoclinic orbits in complex models would support the relevance of our results. The typical cycle of growth-transfer-decay, associated to the homoclinic orbit, has a period of about 6.7 years in our model. The existence and the period of such a cycle in more complex models are also interesting to consider. The period of 6.7 years is comparable to periods of ultra-low-frequency maxima found in previous studies, but it is unclear whether the atmospheric patterns associated with these peaks are of the same type as the patterns in the 10D EOF model.

Other issues arise out of the observed mathematical structures in the model: the role of the 1:3 resonance in the formation of the homoclinic connection isn't resolved yet, nor is the possible link with the existence of a heteroclinic cycle. Such cycles have been observed by Aubry et. al. (1988) in the context of efficient models of a fluid-dynamical problem (see also Holmes et. al. (1997) and references therein). Heteroclinicity is generally associated with symmetries in the model; one can wonder whether atmosphere models like the one that was the subject of this paper are close to such a symmetry, and whether it could be connected with the existence of heteroclinic cycle(s). Finally, since not many examples of bifocal homoclinic bifurcations are known in applications, more numerical details of our example would be of interest.

Continuation of the present work can be envisioned in two directions. Firmer and more rigorous mathematical support for the conjecture of the influence of homoclininc dynamics in atmosphere models is one path to be taken. The other one should head for a clarification of the physical phenomena that are connected to the homoclinic behaviour, and for an understanding of the effect of perturbations (caused by e.g. small-scale atmo-

spheric processes) on this behaviour. We intend to work in both ways.

Chapter 2

Regime transitions and heteroclinic connections in a barotropic atmosphere

Abstract. By interpreting transitions between atmospheric flow regimes as a deterministic rather than a stochastic phenomenon, new insight is gained into the phase-space characteristics of these transitions. The identification of regimes with steady states should be extended with the association of transitions with nearby heteroclinic connections between steady states, as known from the theory of dynamical systems. In the context of a T21 barotropic model of the northern hemisphere, which possesses regime behaviour, steady states are found that correspond with regimes, and heteroclinic connections are approximated using a new algorithm based on adjoint modelling techniques. A 200 year dataset generated by the model is shown to posses spatial preferences in its transitional behaviour that match well with the approximated heteroclinic connections.

2.1   Introduction

The concept of atmospheric circulation regimes is an old idea that is still much studied in meteorology and climate science. The interest in the notion that the atmosphere can be caught, for some time, in particular flow configurations has far from subsided. Apart from the relevance for meteorological studies, the concept of regime behaviour now also draws attention from studies of climate variability and predictability. This is due to the belief that climate variability is related to changes in the probabilities of regime visits (Corti et al. 1998, Palmer 1999). Many studies have been devoted to the question if and where regimes exist in observations and in complex atmosphere/climate model data. Also, the connection of regimes to known (but still ill-understood) climate variability patterns, such as the North-Atlantic Oscillation (NAO), the Arctic Oscillation (AO) and the Pacific/North American (PNA) pattern, is a much debated issue. Surprisingly, the investigation of the dynamical origin of regime behaviour seems to have slowly moved out of sight.

The concept of global bifurcations is an instrument that may enable us to get a better grip on the phenomena of regime behaviour and low-frequency variability and their

underlying dynamical nature. The notion that individual regimes can be associated with separate (simple) attracting structures in the phase space of atmospheric flow exists since Charney and Devore (1979). They identified stable equilibria (steady states, fixed points) with regimes, a hypothesis that was followed and expanded on by e.g. Reinhold and Pierrehumbert (1982), Legras and Ghil (1985), De Swart (1988a,b, 1989) and Itoh and Kimoto (1996). Even though some of these studies speculate on periodic solutions or tori to be the underlying mathematical structures of regimes, the basic idea remains the same: regime behaviour can be identified with temporary stays on or near simple attractors, followed by irregular transitions between these attracting sets.

The multiple attractors have been shown to be amenable to bifurcation analysis by e.g. Legras and Ghil (1985), De Swart (1988a,b, 1989) and Itoh and Kimoto (1996). Continuation and bifurcation software packages such as AUTO (Doedel, 1986) are able to follow steady states and periodic solutions through parameter space. They were used to shed light upon the creation and structure of the simple solutions identified with regimes in low-order atmosphere models. Much more difficult however are the mechanisms of transitions between these solutions. That may be the reason that the, from a dynamical systems point of view most natural possibility of *global bifurcations*, i.e. of deterministic connections being created between different solutions (or from a solution to itself), hasn't gotten much attention yet. The most common transition mechanism invoked is stochasticity: transitions induced by stochastic perturbations in the model equations. Introducing such perturbations is usually motivated by referring to physics and dynamics (e.g. small-scale, high frequency motions or convection) that are not explicitly resolved in the relatively simple models used for bifurcation studies in relation to regime transitions. However, introducing stochastic terms or interpreting small-scale, high-frequency behaviour in complex, deterministic atmosphere models as noise is not entirely satisfying. Ignoring the deterministic origin of such stochasticity may obscure relevant information about regime behaviour. In this paper we will hold on to this deterministic nature, and use the notion of heteroclinic connections to get a better understanding of regime transitions.

Heteroclinic orbits are solutions of (deterministic) model equations that connect one invariant set, e.g. a steady state solution, with another. Homoclinic orbits connect an invariant set with itself. Hetero- resp. homoclinic bifurcations are bifurcations in which such orbits, or connections, are created. They are mentioned a few times in the literature as a possible mechanism for transitions between regimes. Legras and Ghil (1985) use a 25-dimensional equivalent-barotropic model, perform a bifurcation analysis of its steady states and find, by numerical integration, transitions between regimes. Heteroclinic orbits as a possible explanation are mentioned, but not expanded on. In De Swart (1988a,b, 1989) a 6-component spectral model was found to possess multiple attractors, some of them with chaotic windows (parameter intervals with chaotic behaviour), that could be identified with regimes. However, transitions between these attractors (vacillation) were not observed. By adding stochastic perturbations, a transition mechanism was created (De Swart and Grasman, 1987). Also, by increasing the number of spectral components to 10, the model could be brought to vacillatory behaviour, without stochastic perturbations. But no attempt was made to find heteroclinic orbits. Itoh and Kimoto (1996,1997,1999) perform bifurcation analyses of a two-layer T15 model and of a five-layer T21 model. They find multiple attractors as well as wandering between different attractor ruins, and

call this wandering an example of "chaotic itinerancy". It is not made clear if and how this chaotic itinerancy is related to the concept of homo- and heteroclinic connections.

The notion that regime transitions are not entirely random is recognised in, amongst others, Mo and Ghil (1988), Kimoto and Ghil (1993b) and Itoh and Kimoto (1996, 1997, 1999). These studies identify more than two regimes and point out that some transitions between regimes are more probable than others. Using this statistical approach, cycles of most probable transitions are identified; these are called "preferred transitions". Itoh and Kimoto (1997) refer to the notion of "explosive bifurcations", but they don't try to locate heteroclinic connections, nor do they mention them. In the present study, something different is meant with the notion "preferred transitions": it refers to the idea that transitions between regimes tend to follow certain routes through phase space; routes that can be associated with approximate heteroclinic connections. The focus is on preferences of the behaviour *during* the transitional episodes, not on the most probable order of regime visits without looking at characteristics of the transition phases, as in Mo and Ghil (1988), Kimoto and Ghil (1993b) and Itoh and Kimoto (1997).

The study by Plaut and Vautard (1994) shows, performing Multichannel Singular Spectrum Analysis (MSSA) on observational data, that regime behaviour is connected to the existence of low-frequency oscillations (LFOs) in the atmosphere. LFOs influence regime behaviour but not in a systematic way; rather, the interaction between two Atlantic sector LFOs (one of 70 days and one of 30-35 days) seems to be one mechanism, among others, that can produce high-amplitude anomalies (regimes). It remains unclear, though, whether these anomalies correspond to (quasi-) stationary states (thus, if the system evolution slows down). The authors hypothesize that LFOs should be associated with the presence of unstable periodic orbits in phase-space. An important conclusion is that regime transitions, at least in the Atlantic sector, are not produced randomly by synoptic transients. This supports the choice for the deterministic (instead of stochastic) point of view taken in the current paper.

In Crommelin (2002), the appearance of homoclinic dynamics in an atmosphere model gets more detailed attention. In this paper, a two-layer model is analysed which uses Empirical Orthogonal Functions (EOFs) instead of spherical harmonics as its basis functions (see Achatz and Branstator (1999) for the derivation and formulation of the model). With 10 EOFs, some realism was still retained and at the same time it was possible to find evidence for the occurrence of homoclinic dynamics. The dynamics on long timescales turned out to be dominated by a homoclinic orbit of the *bifocal* type: an orbit attached to a steady state with two complex pairs as its leading eigenvalues. A glimpse of a heteroclinic connection between two equilibria was also found, but the role of one of the equilibria could only be made visible for small parameter regions. The other equilibrium, very close to the climatic mean state, is prominent in the model behaviour over large parameter ranges. The behaviour was therefore identified as driven by a homoclinic, not a heteroclinic connection.

In the fluid dynamics community, the possibility of heteroclinic connections as an explanation for transitions between steady states has been recognised before. Examples were found in e.g. Rayleigh-Benard convection (Proctor and Jones, 1988) and the turbulent boundary layer (Aubry et al. (1988), Armbruster et al. (1988), Holmes et al. (1996,1997)). Knobloch and Moehlis (2000) provide an overview of mechanisms that produce bursting

in hydrodynamical systems (sudden transitions from and to regular, repeatedly realized states); again, heteroclinic connections play an important role. They interpret transitions between different types of behaviour as a form of bursting. Although no examples are mentioned from geophysical fluid dynamics, such a concept of bursting could equally well apply to the case of regime-transitions in atmospheric dynamics. The methods and concepts used in the study of these fluid-dynamical examples could be very useful in the investigation of regime behaviour and low-frequency variability of the atmosphere. This paper is partly inspired by these examples.

The order of things to come in this paper is as follows. In section 2 the barotropic model that was used will be described. The model shows bimodality and regime behaviour, treated in section 3. The regimes can be associated with steady states; this is described in section 4. In section 5 the role of heteroclinic connections will be explained. A new algorithm, based on adjoint modelling techniques, to calculate approximations of such connections will also be presented in this section, together with its results. Section 6 contains a discussion and summary.

2.2   A barotropic model

For this study a barotropic spectral model was used. It is the same model that was used in Selten (1995), a standard model with realistic orography ("mountains") and a forcing that was calculated from observations in order to get a realistic climate mean state and realistic low-frequency variability.

The model is based on the barotropic vorticity equation, and reads:

$$\frac{\partial}{\partial t}\xi = -\mathcal{J}(\psi, \xi + f + h) + k_1\xi + k_2\Delta^3\xi + \xi^* \tag{2.1}$$

This partial differential equation describes the temporal evolution of the relative vorticity field $\xi(t, \lambda, \phi)$, a scalar field, on the sphere. The spatial coordinates are the longitude λ and latitude ϕ. The streamfunction $\psi(t, \lambda, \phi)$ is related to the vorticity via $\xi = \Delta\psi (= \nabla^2\psi)$. The terms multiplied by k_1 and k_2 are damping terms (Ekman friction resp. scale-selective damping); f denotes the Coriolis parameter and h is the (nondimensional) orographic height. $\mathcal{J}$ denotes the Jacobi operator, which reads $\mathcal{J}(A, B) = \frac{1}{\cos\phi}\left(\frac{\partial A}{\partial \lambda}\frac{\partial B}{\partial \phi} - \frac{\partial A}{\partial \phi}\frac{\partial B}{\partial \lambda}\right)$. The (time-independent) vorticity forcing ξ^* is based on observations:

$$\xi^* = \mathcal{J}(\psi_{\rm cl}, \xi_{\rm cl} + f + h) - k_1\xi_{\rm cl} - k_2\Delta^3\xi_{\rm cl} + \overline{\mathcal{J}(\psi', \xi')}. \tag{2.2}$$

This procedure was proposed by Roads (1987). It involves the use of the observed climate mean states at 500 hPa of the winter season $\xi_{\rm cl}$ and $\psi_{\rm cl}$. The last contribution to the forcing, due to the transient eddy forcing, is calculated using (winter) anomalies of the 10-day running mean at 500 hPa, ξ' and ψ'. The overbar denotes, as usual, time average.

Equation (2.1) is projected onto spherical harmonics and triangularly truncated at wavenumber 21 (i.e. T21), yielding a coupled set of nonlinear ordinary differential equations for the time-dependent spectral coefficients. By discarding all modes with zonal wavenumber plus total wavenumber being even, a model of hemispheric flow is obtained
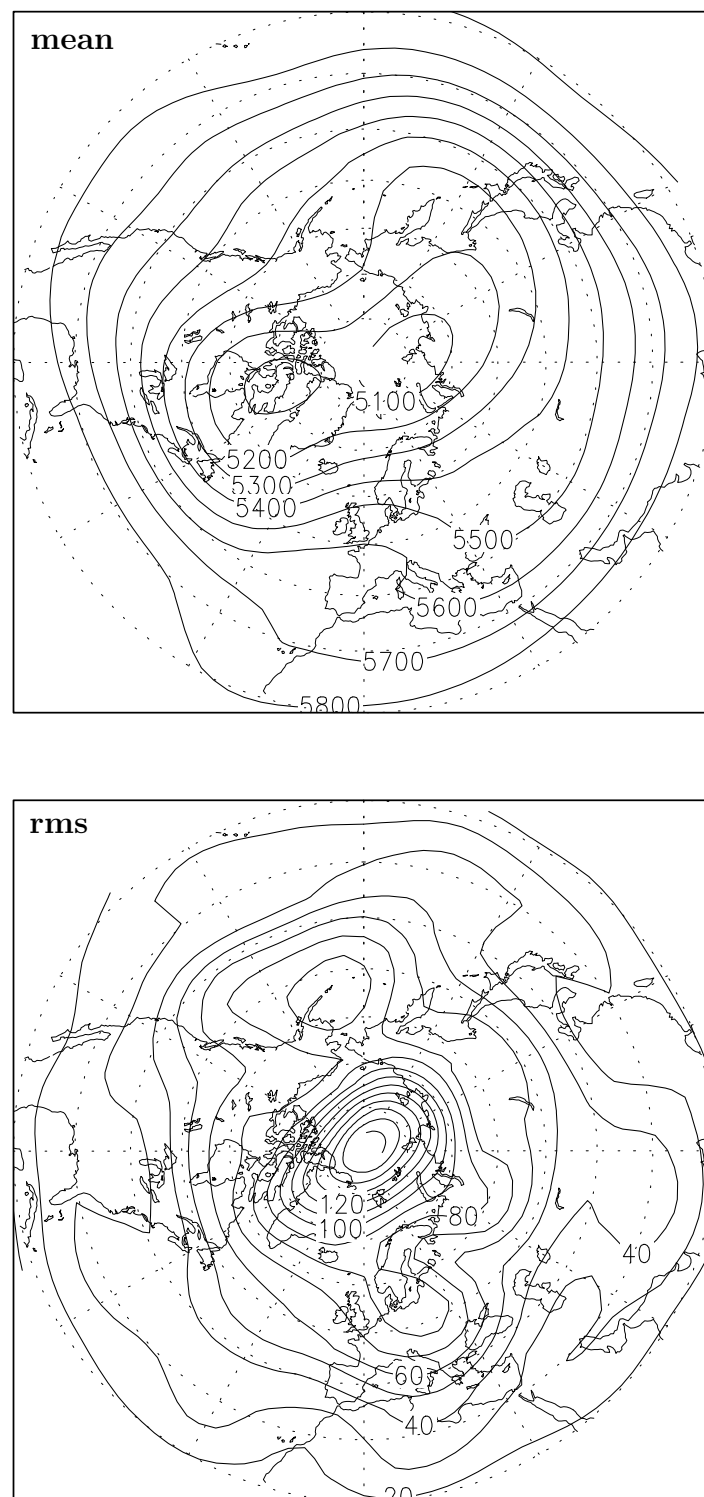
Figure 2.1: *Top: time mean of the 500 hPa geopotential height Z_{500} as produced by the T21 barotropic model. Contours are drawn every 100 meter. Bottom: Z_{500} rms of the T21 model. Contour interval is 10 meter.*

with a total number of 231 variables. It will serve to describe the (barotropic) atmosphere of the northern hemisphere. For more detailed information on the formulation of the model the reader is referred to Selten (1995).

The model has been integrated for 73 000 days (200 years), using a Runge-Kutta fourth-order routine with a 30 minutes timestep and daily output. The initial state was a realistic flow field, so no spin-up time was needed. The timescale of the Ekman damping was set to 15 days; the scale-selective damping was tuned such that the damping timescale is 3 days for wavenumber 21. Figure 2.1 contains the time-mean and variability (root mean square, or rms) of the thus obtained 200 year dataset. Shown are the mean and rms of the 500 hPa geopotential height Z_{500}, which is calculated from the streamfunction (at 500 hPa) using the so-called linear balance equation $g_0 \nabla^2 Z_{500} = \nabla \cdot (f \nabla \psi)$, $g_0 = 9.80665 \, \mathrm{m \, s^{-2}}$ (see Holton, 1992). The mean resembles the mean field of the real atmosphere quite closely. The rms is not as close to the observations, but still quite realistic considering the model resolution. Most notable are the unrealisticly high variability above the north pole and the location of the rms-maximum over Europe, which should be over the north-east Atlantic. For a more detailed comparison with observations, see Selten (1995). The 200 year dataset will be analysed in order to get a better understanding of its low-frequency aspects.

2.3   Bimodality

EOFs of the streamfunction data are calculated; in figure 2.2 the streamfunction patterns corresponding to the leading four EOFs are shown. For the calculation, the time mean state was subtracted from the data; the norm chosen was the kinetic energy norm (see appendix 2.A). The amount of kinetic energy variance represented by EOFs 1–4 is 18.3, 9.8, 6.8 and 6.0 percent, respectively; together these four represent 40.9 %. The principal components (PCs) were calculated by projecting the (spectral) data onto the EOFs.

The shape of the two-dimensional Probability Density Function (PDF), made from the projection of the full 73000 days dataset onto the PC1,PC2–plane, reveals a bimodal structure with a main maximum at PC1 $\approx$ 0.0055 (MAX1) and a second maximum at PC1 $\approx$ −0.0089 (MAX2). The PDFs made from projections onto the PC2,PC3–, the PC2,PC4– and the PC3,PC4–plane (not shown) yielded unimodal distributions; the bimodal structure mainly projects onto EOF1. Figure 2.3 (dashed lines) shows the PDF in the PC1,PC2-plane; the various symbols in the figure will be explained later. The PDFs were calculated by collecting the projected datapoints in 15 × 15 bins and calculating the corresponding histogram.

The locations of the two maxima of the two-dimensional PDF were estimated in all 231 PCs by first identifying the PC1 values of the two maxima. Then, two sets of datapoints were created by selecting all datapoints with their PC1 value in a small band around the two maxima. For all higher PCs two one-dimensional PDFs and their maxima were calculated, thus giving estimates for the location of the two maxima in all PCs. The Z_{500} patterns corresponding to the maxima are shown in figure 2.4 (note that these patterns consist of contributions of all 231 PCs). The distances to these two maxima during the model run were calculated, using again the kinetic energy norm. For a segment of the
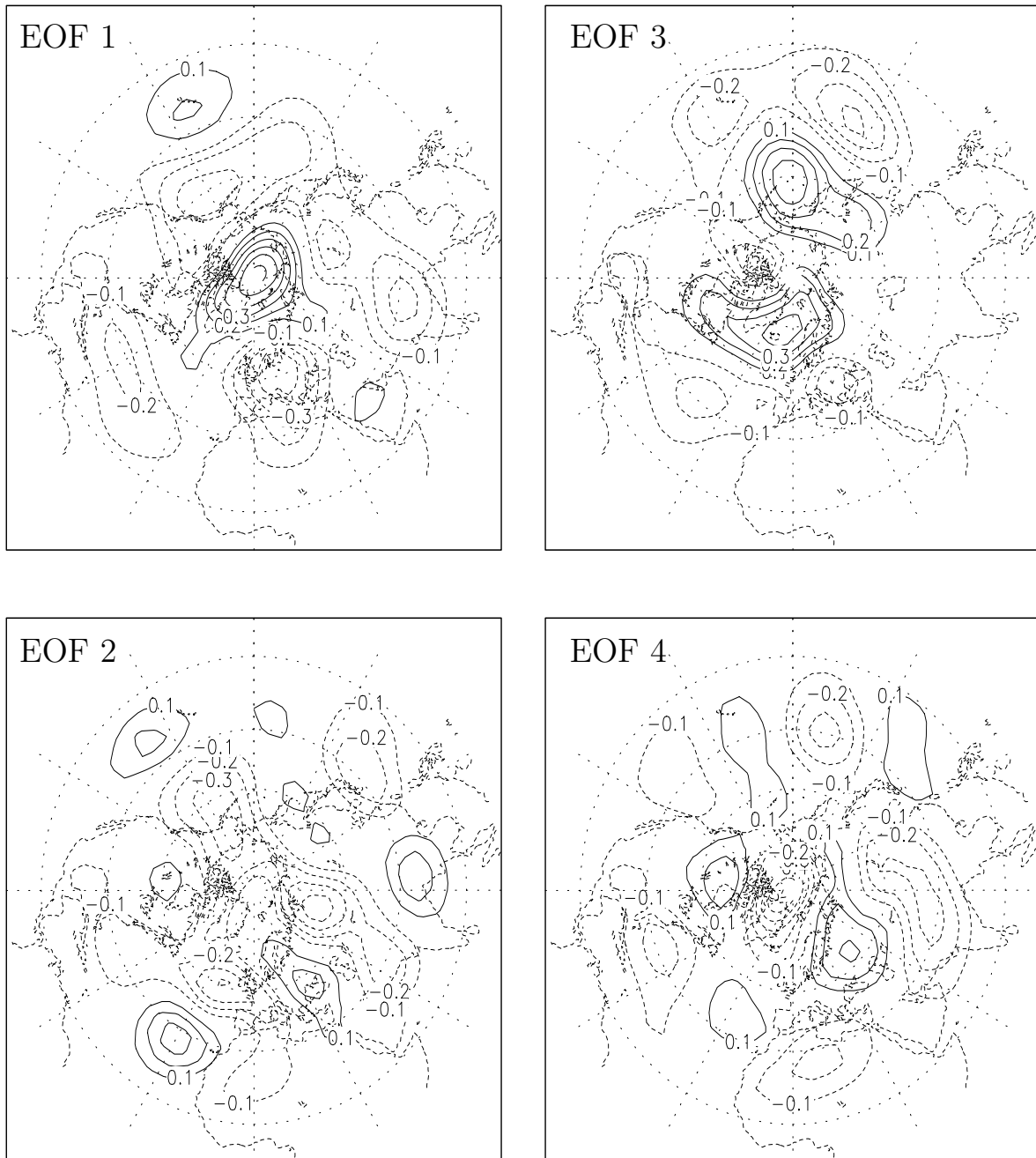
Figure 2.2: *EOFs 1-4 calculated from the 200 year dataset produced by the T21 barotropic model. The kinetic energy norm was used for the calculation.*
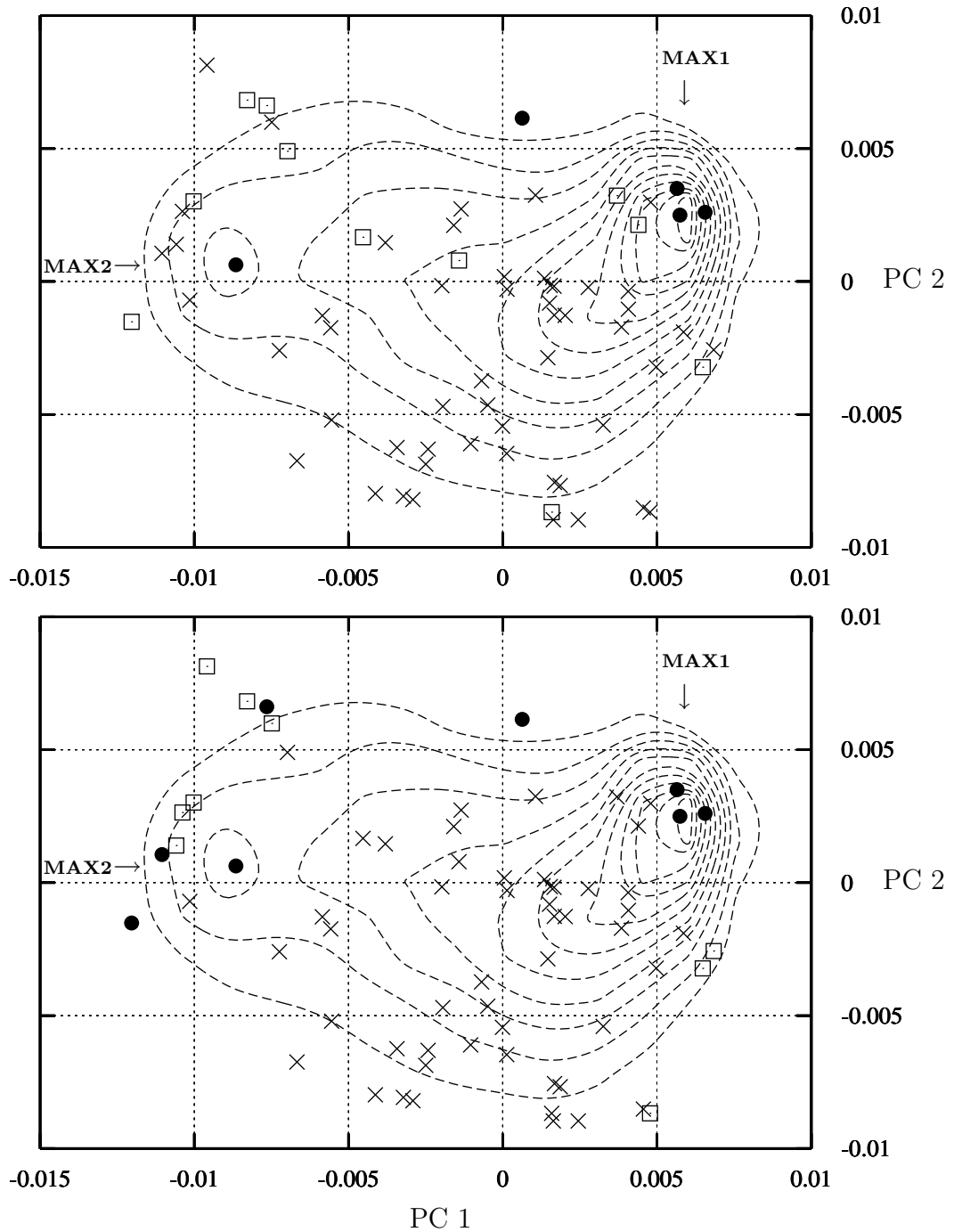
Figure 2.3: *PDF and steady states in PC1,PC2–plane. The dashed lines are the contours of the PDF (contour distance: 200 datapoints per bin, number of bins: 15×15). Top: steady states with number of unstable eigenvalues $n_u \leq 6$ ($\bullet$); $7 \leq n_u \leq 10$ ($\boxdot$); $n_u > 10$ ($\times$). Bottom: steady states with ratio of projection onto the leading 2 EOFs $R_2 > 0.5$ ($\bullet$); $0.35 < R_2 < 0.5$ ($\boxdot$); $R_2 < 0.35$ ($\times$).*
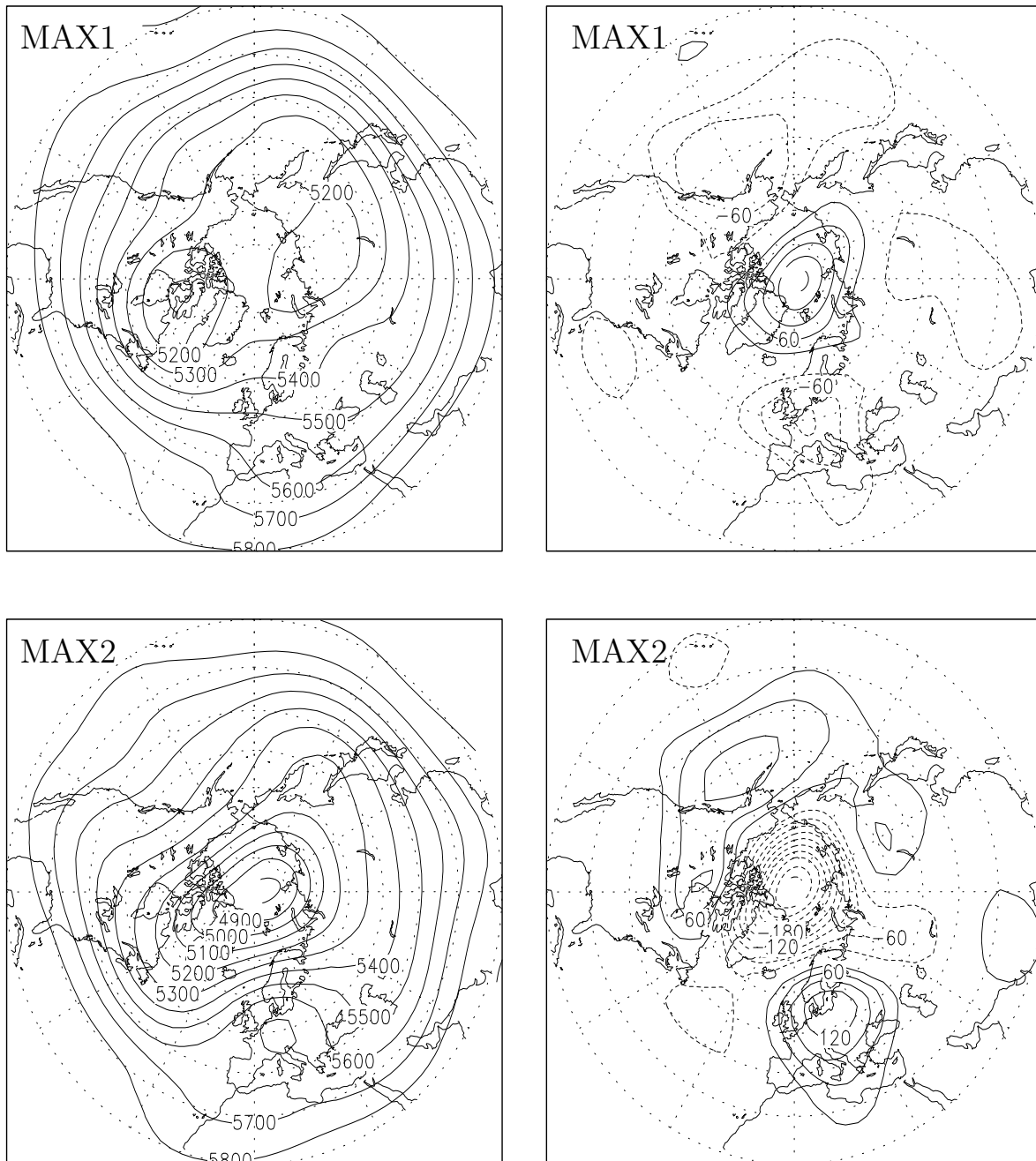
Figure 2.4: *Patterns corresponding to the maxima of the PDF. Top: MAX1, bottom: MAX2 (see text). Shown left are patterns of the 500 hPa geopotential height field Z_{500} (contour interval 100 m). On the right are the corresponding Z_{500} anomalies (contour interval 30 m).*

data, the timeseries of these distances are shown in figure 2.5. They show persistent periods close to the maxima as well as rapid transitions between them.

The PDF-maxima MAX1 and MAX2 can be interpreted as flow regimes, the transitions between them as vacillatory behaviour. The flow-pattern (figure 2.4) of MAX1 represents a situation of intensified zonal flow, the pattern of MAX2 shows a blocked flow over Europe (a split jetstream and a positive anomaly, or "high", over Europe). Both regimes have previously been found to correspond with steady states in low-order models (Charney and DeVore (1979), Legras and Ghil (1985)). With this correspondence in mind, one might expect the presence of (saddle) equilibria at or near MAX1 and MAX2 in our T21 model as well. In the next section, the presence of equilibria in the model will be investigated. The clear transitions between MAX1 and MAX2 suggest the influence of two heteroclinic connections between these assumed equilibria. After all, the model is completely deterministic; it does not contain any stochastic process that could have induced the transitions. Moreover, from the swiftness of the transitions and the persistence of the regimes it can be seen that the system evolution slows down near MAX1 and MAX2. The possibility of unstable periodic orbits being present in phase space that pass both MAX1 and MAX2 still leaves the question why the motion slows down near these maxima. The influence of heteroclinic connections can provide an explanantion, since such connections always end with a slowing down of the system evolution, due to the approach of a fixed point at the end of each connection.

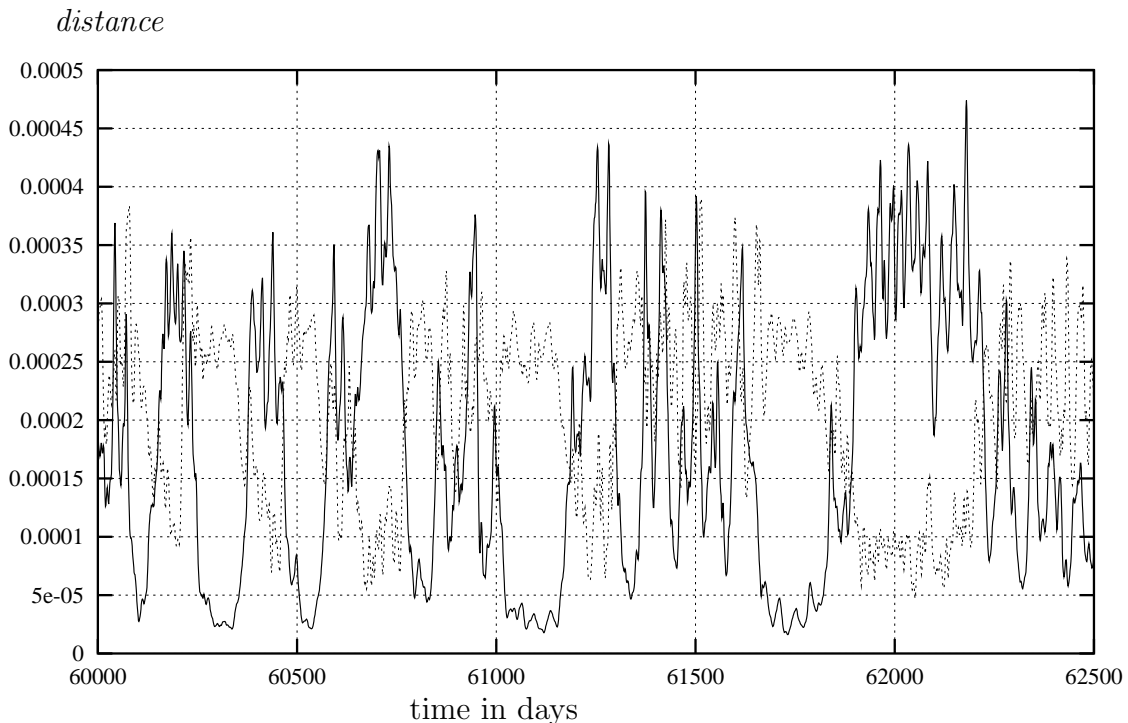The difference between MAX1 and MAX2 is largely spanned by EOF1, which resem-

Figure 2.5: *Phase space distances to two maxima MAX1 (solid) and MAX2 (dotted) during a data segment of 2500 days. Distances are given as differences in turbulent kinetic energy; horizontal axis denotes time in days.*

bles the AO pattern (Thompson and Wallace, 1998). From a phase space point of view, the AO pattern in this model seems to be the direction, or phase space vector, connecting the two regimes. The blocking-character of the MAX2-regime is a regional aspect of a hemispheric pattern that resembles one of the phases of the AO. Long timescale variability of the AO should be related to the dynamics created by, or associated with, the near-presence (in parameter space) of a heteroclinic cycle (connection back and forth). In this perspective, variability of the AO has its dynamical origin in the (ultra-)low-frequency aspects of the dynamics of regime transitions. A better understanding of the dynamical mechanism(s) of regime transitions will help to better understand climate variability patterns such as the AO.

2.4 Equilibria

For models like the one under investigation there are two different (numerical) techniques available to find equilibrium solutions (fixed points). One is continuation: starting from a known steady state at specific (often unrealistic) parameter values, it is in principle possible to follow this steady state as parameter values are changed, toward realistic values. For an atmosphere model, one may think of starting at zero flow in the absence of forcing, and gradually turning on the forcing to a realistic intensity. A drawback of the continuation method is that it is hard, sometimes impossible, to find different (disconnected) branches of fixed points. Using continuation one easily (if not: generally) misses equilibria.

The other, less elegant method, used in the current study, is a way of "brute force calculation": bluntly calculating zeroes of the set of model-ODEs. For low-order models one can think of a Newton-like rootfinding algorithm; for larger models some sort of minimization-technique can be used (direct rootfinding in large models often doesn't work because of the presence of many local minima in phase space that are not solutions). For this study, a method was chosen that minimizes a scalar function, given its gradient. It was used before by Branstator and Opsteegh (1989). The squared norm of the vorticity-tendency $\partial \xi / \partial t$ served as the scalar function (denoted F); it is clear that a zero of this function is both a necessary and sufficient condition for the existence of a fixed point. The precise definition of F and the derivation of the gradient of F with respect to the model variables can be found in appendix 2.B. The algorithm used comes from the NAG software library (routine E04UCF); it is a quasi-Newton method.

Many different states, sampled from all over the model attractor, need to be used as initial guesses for the minimization routine, in order to find as many different steady states as possible. This was done by starting the minimization from 5000 datapoints taken from the 200 year model integration, each 10 days apart. This set of 5000 points is assumed to give a reasonable coverage of all regions of the attractor. Not all 5000 minimizations converged to points with acceptably small F: about 1000 converged to points with $F \leq 10^{-14}$, giving 68 different fixed points in total. For comparison: the values of F before minimization were of order 10^{-4}.

It must be stressed that we do not claim to have found all the equilibrium solutions

of the model. However, it appears we have found the most relevant equilibria, namely those fixed points located in or near the two maxima of the PDF. We will apply two (dependent) criteria to assess the importance of the various fixed points for the large-scale, low-frequency dynamics of the barotropic model. One is the number n_u of unstable eigenvalues of the fixed points: fewer unstable eigenvalues suggests that the equilibrium is "more attracting" and therefore more likely to attract and influence the system as it evolves in time. We have calculated the eigenvalue-spectrum of each fixed point by evaluating the Jacobian matrix (see Appendix 2.B) and calculating its eigenvalues. The result is shown in figures 2.6 and 2.3, which will be discussed later.

If the fixed points with the least unstable eigenvalues are indeed most likely to influence the dynamics, they can be expected to project strongly onto the leading EOFs. This should be visible using the second criterion: the extent to which each equilibrium state projects onto the leading EOFs. The behaviour of the dominant EOFs is most likely to be influenced by equilibria that strongly project onto them. Fixed points that lie far outside the subspace of the leading EOFs, and hence project relatively strongly onto the trailing EOFs, are less likely to be very relevant for the large scale dynamics of the model. Projecting each equilibrium onto the EOFs gives a vector $(a_1, a_2, \ldots, a_N)^{\mathrm{T}}$ of coefficients a_i. The total number of EOFs, N, equals 231. By calculating the ratio

$$R_M = \sum_{i=1}^{M} a_i^2 \Big/ \sum_{i=1}^{N} a_i^2 \qquad (2.3)$$

for low values of M, one gets an impression of the contribution of the dominant EOFs to the equilibrium flow patterns. In fact, since the kinetic energy norm was used in the calculation of the EOFs, R_M equals the fraction of the turbulent energy contained in the first M EOFs.
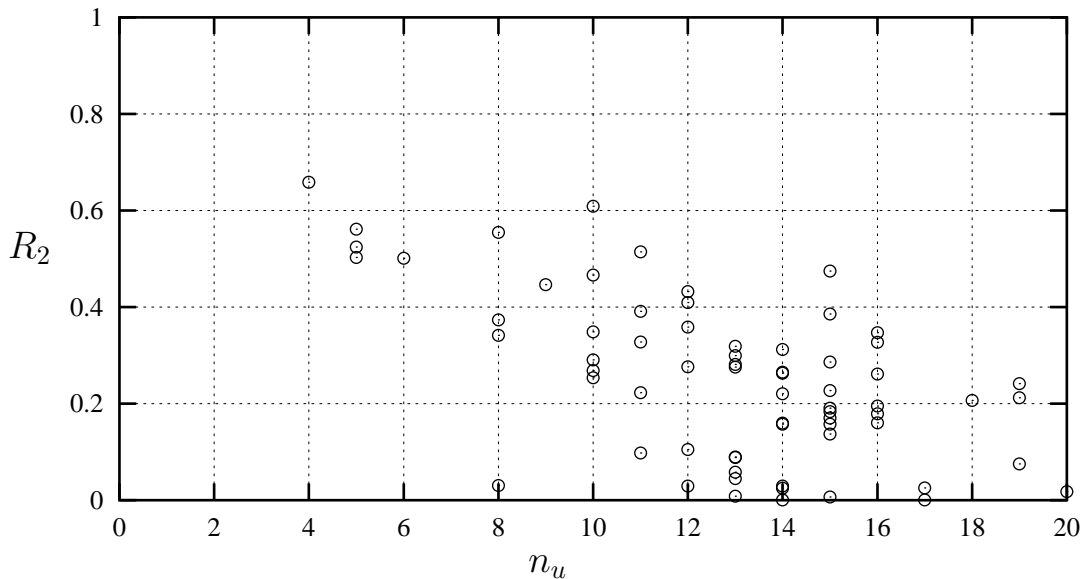

Figure 2.6: *Number of unstable eigenvalues n_u versus ratio R_2 of projection onto the leading 2 EOFs (see text) for each steady state of the barotropic model.*

Figures 2.6 and 2.3 contain the results of the calculations. Figure 2.6 shows a scatter plot of R_2 versus n_u, the number of unstable eigenvalues. There is a clear correlation between the two, which confirms the intuitive picture described earlier: the lower n_u, the more the equilibrium influences the dynamics of the system and hence the stronger it projects onto the leading EOFs (as was pointed out by one of the reviewers, figure 2.6 is in fact a test of the validity of applying the first criterion). Figure 2.3 shows the location of all the fixed points when projected onto the PC1,PC2 plane. Different symbols have been used in order to show the locations of the equilibria with low n_u (top) and high R_2 (bottom). Also shown are the contour lines (dashed) of the PDF. Applying the two criteria, low n_u and high R_2, gives a selection of five steady states: the ones with $n_u \leq 6$ (they all belong to the group with highest R_2 as well). One of the equilibria is located near MAX2, the second PDF-maximum; three are near the first maximum MAX1. The hypothesis that these maxima are related to the near-presence of relatively stable fixed points is thus supported. Note that the instability of the fixed points, although weak, makes it unlikely for them to be located exactly in the maxima of the PDF of the entire 231-dimensional phase space (after all, the system can never be exactly at an unstable fixed point). The fifth steady state lies rather eccentric to the PDF. We do not know why this is so.

2.5 Heteroclinic connections

The transitions between atmospheric flow regimes are usually thought of as stochastically induced phenomena. However, taking into consideration that in our model flow regimes can indeed be associated with steady states and that the model doesn't contain stochastic terms, it is, from a dynamical systems point of view, more natural to think of regime transitions as manifestations of heteroclinic behaviour. We hypothesize that the model is not very far (in parameter-space) from a situation in which there is one (or several) phase-space orbit(s) going exactly from a steady state associated with one regime to a steady state associated with another regime, and another orbit going back. Although such a situation requires the tuning of several parameters, and is thus unlikely to be exactly realised with the parameter setting that was chosen for the model, we expect that the influence of the connections is still felt by the system and manifests itself as transitions between regimes. In this section we will first describe in more mathematical detail what a heteroclinic connection is, and what the bifurcation scenario could be in which these connections are broken but still play a role. Then we both present evidence to support the hypothesis and show that the regime transitions in the model have a tendency to follow paths that are approximations of heteroclinic connections.

2.5.1 A bifurcation scenario

A heteroclinic connection, or heteroclinic orbit, is the phase space projection of a solution of the model equations that connects one invariant set with another. The invariant sets can be steady states, periodic solutions, tori, etc. In what follows we will only consider heteroclinic connections between steady states.

Let us put these notions in more formal terms. Let $\dot{y} = G(y, \alpha)$ be a set of ordinary differential equations with parameter-set α. A heteroclinic connection $\tilde{y}(t)$ is the projection onto phase space of a solution that asymptotically approaches in forward and backward time two different steady states y_- and y_+:

$$\lim_{t \to +\infty} \tilde{y}(t) = y_+, \quad \lim_{t \to -\infty} \tilde{y}(t) = y_- \tag{2.4}$$

A set of heteroclinic connections forming a closed loop is called a heteroclinic cycle; a heteroclinic connection going from a steady state back to itself (i.e. $y_- = y_+$) is called a homoclinic orbit. For general background information on heteroclinic orbits, see Kuznetsov (1995). A detailed review of the theory of heteroclinic cycles is provided in Krupa (1997).

Our hypothesis is that regime behaviour, or vacillation, is eventually generated by heteroclinic connections going back and forth between invariant sets which are associated with regimes (i.e. a simple heteroclinic cycle). As previously mentioned, in this study we will assume the invariant sets to be steady states. The presence of this structure of two connections is not generic; in general, one or more parameters must be tuned in order to create the two connections. It is difficult to estimate a priori the number of parameters that must be tuned (the codimension). However, looking at the dimensions of the involved stable and unstable manifolds can give us some information. We choose the steady state near MAX1 with 4 unstable eigenvalues as the first fixed point y_1 and the steady state near MAX2, which has 6 unstable eigenvalues, as the second fixed point y_2. Both fixed points are hyperbolic. $W^u(y_i)$ denotes the unstable manifold of y_i, $W^s(y_i)$ the stable manifold. A single heteroclinic connection from y_1 to y_2 must lie entirely in both $W^u(y_1)$ and $W^s(y_2)$. Since $\dim(W^u(y_1)) = 4$ and $\dim(W^s(y_2)) = 225$, and since in general 3 parameters must be tuned in order for a 4-dimensional (4-d) manifold and a 225-d manifold to have a 1-d intersection in a 231-d space , the creation of the 1-d intersection of $W^u(y_1)$ and $W^s(y_2)$ in the 231-d phase-space has codimension 3.

To formulate the argument in more precise terms, we need some definitions, taken mainly from the book by Guckenheimer and Holmes (1983). First of all, the *transversality theorem* tells us that the intersection of two manifolds of dimensions m_1 and m_2 in an n-dimensional space has, generically, dimension $m_1 + m_2 - n$. If $m_1 + m_2 < n$ there is in general no intersection. Second, a *transversal intersection* of two manifolds in an n-dimensional space is an intersection in which the tangent spaces of the two manifolds in the intersection point(s) span the total n-dimensional space. A transversal intersection is persistent under small perturbations of the two manifolds. Finally, the codimension of a bifurcation can be defined as the smallest dimension of a parameter space which contains the bifurcation in a persistent (structurally stable) way. From these definitions follows that the creation of a 1-dimensional intersection of two manifolds M_1 and M_2 of dimensions m_1 and m_2 in an n-dimensional space with $m_1 + m_2 < n$ (which is a nontransversal intersection) has codimension $p = n + 1 - m_1 - m_2$, since in the space $\mathbb{R}^n \times \mathbb{R}^p$ (variables plus parameters) the two manifolds M_1 and M_2 (now of dimensions $m_1 + p$ and $m_2 + p$) have in general a transversal intersection of dimension $(m_1 + p) + (m_2 + p) - (n + p) = 1$. Applying this to the situation at hand shows that the 1-d intersection of a 4-d manifold with a 225-d manifold in a 231-d space happens in a persistent way in a parameter space of (at minimum) 3 dimensions. Thus, the codimension of the bifurcation in which this

intersection is created is 3. For a more detailed treatment of these issues the reader is referred to Guckenheimer and Holmes (1983) and Kuznetsov (1995).

Unfortunately, other invariant objects than the two fixed points can make the situation more complicated. $W^u(y_1)$ and $W^s(y_2)$ can also intersect with the stable and unstable manifolds of e.g. another fixed point, or periodic orbit. As for the reverse connection, from y_2 to y_1: $W^u(y_2)$ and $W^s(y_1)$ generically have a 2-d (transversal) intersection, so this reverse heteroclinic connection at least seems to be more generic than the y_1 to y_2 connection. The dimensionality of $W^u(y_2)$ and $W^s(y_1)$ would even allow for a 2-d connecting surface rather than a single, 1-d heteroclinic connection. But again, other invariant objects can complicate the situation.

In other physical systems in which heteroclinic cycles play an important role in the dynamics, the existence of such cycles is usually connected to the presence of symmetries in the system. Symmetries induce the presence of invariant subspaces in phase-space, which can make the existence of a heteroclinic cycle generic, without needing to tune parameters (as is needed to create a heteroclinic cycle in a nonsymmetric system). Such a cycle will be robust for perturbations that preserve the symmetry, as these perturbations leave the invariant subspace structure intact. See Krupa (1997) for a detailed account. Fluid-dynamical examples of heteroclinic cycles have been found in e.g. Rayleigh-Benard convection (Proctor and Jones, 1988) and the turbulent boundary layer (Holmes et. al. (1996, 1997) and references therein). At this stage we can only speculate about the (near-) presence of symmetries in the barotropic model, and their possible role in the formation of a heteroclinic cycle. In this context, it is interesting to note that the two PDF-maxima (and their corresponding steady states) are located roughly symmetric around the time mean state (where all PCs are zero). This can also be seen in the right panels of figure 2.4: the anomalies corresponding to MAX1 and MAX2 have roughly the same structure, but opposite sign. Another source of symmetry could be the dominant wavenumber 2 contribution to the orography. The question of symmetry, although impossible to answer at this moment, is of interest because a symmetry could be the common source of regime behaviour in different atmosphere models (e.g., if a certain (idealised) orography implies a symmetry that favors the creation of a heteroclinic cycle corresponding to regime transitions, it could explain the occurrence of regime behaviour in various atmosphere models with similar orography).

As mentioned previously, we hypothesize a point to be present in parameter space where a heteroclinic cycle exists between two fixed points corresponding to different regimes. Moving away from this bifurcation point (to more realistic physical conditions) one, or both, of the heteroclinic connections is almost certainly broken. If indeed a symmetry is related to the existence of the cycle, leaving the bifurcation point will probably perturb the symmetry (forced symmetry breaking). The phase-space structure that is left is expected to still bear traces of the former cycle, since it generates regime behaviour in the system. The structure should somehow account for the asymmetry of the regime behaviour, the fact that the zonal regime is visited more often than the blocked regime – the PDF is bimodal but not symmetric (in other studies, e.g. Itoh and Kimoto (1999), asymmetric PDFs have also been found). A possible explanation is that a homoclinic orbit connected to the zonal fixed point could be created when the cycle is broken. This would explain the findings in Crommelin (2002) where such a homoclinic orbit was detected. In

that study, the possibility of heteroclinic behaviour was also mentioned. The behaviour of the current T21 model was very briefly studied there, and related to homoclinic dynamics; the heteroclinic influence was not detected due to the briefness of the inspection of the T21 model there.

It is known that the multiple equilibria associated with regime behaviour are the result of orographically induced saddle-node bifurcations that make the sheet of fixed points "fold" in parameter space, see Charney and DeVore (1979). Furthermore, Hopf-bifurcations due to barotropic instability occur naturally in barotropic models, see e.g. Legras and Ghil (1985) and De Swart (1988a, 1988b, 1989). Among the phenomena involved in the simultaneous occurrence of a Hopf- and a saddle-node bifurcation, a so-called fold-Hopf (or zero-Hopf) bifurcation, are heteroclinic connections, and homoclinic orbits resulting from perturbations, see Kuznetsov (1995). Thus, a scenario is possible in which orography and barotropic instability result in regime behaviour which originates from a heteroclinic cycle and which is deformed into more homoclinic-type behaviour, favoring the zonal regime, due to forced symmetry breaking. We hope to report on this bifurcation scenario in a future paper. Let us furthermore mention that homoclinic bifurcations produce periodic orbits; in the cases of a Shilnikov-type or a bifocal-type homoclinic bifurcation an infinite number of (unstable) periodic orbits are generated. This could be related to low-frequency oscillations found by Plaut and Vautard (1994). However, much more study will be required to sort this out.

The regime behaviour in the barotropic model is clearly visible. Moreover, the regimes were found to correspond with nearby steady states. Having the above bifurcation scenario in mind, this means that the model cannot be too far from the point in parameter space where the heteroclinic cycle exists. The heteroclinic connections are most likely broken, but there will still be orbits that approximate the former connections and thus go from the vicinity of one steady state to the vicinity of the other. These approximations are what we will focus at in the next section. We will not attempt to prove the validity of the bifurcation scenario just sketched in the context of the barotropic model; such an attempt must start with the investigation of low-order models and of possible symmetries.

2.5.2   Numerical approximation

It is known that a heteroclinic connection from one steady state A to another B should start in the unstable manifold $W^u(A)$ of state A and end in the stable manifold $W^s(B)$ of state B. Furthermore, it takes infinitely long to follow the entire connection, from its starting point A to its end point B. In practice a heteroclinic connection is therefore approximated as a (finite time) connection starting in the linear approximation $E^u(A)$ of $W^u(A)$ at small but finite distance ε_A from point A and ending at distance ε_B from point B in the linear approximation $E^s(B)$ of $W^s(B)$ (with $0 < \varepsilon_{A,B} \ll 1$). A more detailed account is given in e.g. Doedel and Friedman (1989).

As said, it is unlikely that we are exactly in the bifurcation point and hence unlikely that exact heteroclinic connections exist; however, we can expect that there are still phase space orbits running from the immediate vicinity of one fixed point to the immediate vicinity of the other. To verify this, we took the most stable fixed point ($n_u = 4$) near the main PDF maximum (MAX1) as point A and the (only) fixed point near the submaximum

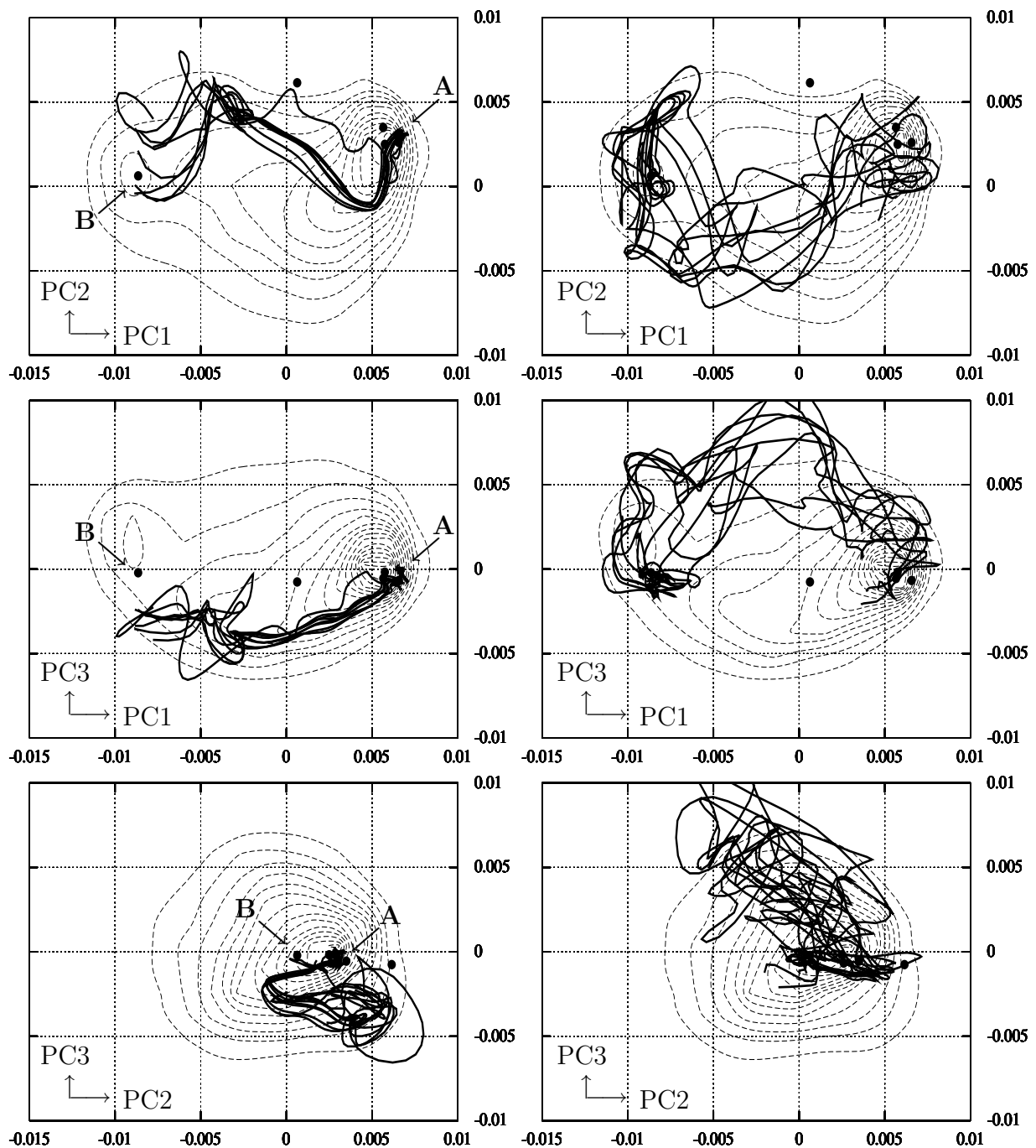Figure 2.7: *Projection of transition orbits onto planes of PC1,PC2 (top), PC1,PC3 (middle) and PC2,PC3 (bottom). Left: orbits starting close to steady state A near MAX1. Right: orbits starting close to steady state B near MAX2. Also shown are the contourlines (dashed) of the PDF of the 200 year dataset and the locations of the 5 selected steady states (see text). The locations of A and B are indicated in the left panels.*

(MAX2) as point B, and performed 2 series of 500 integrations of 100 days each. The integrations of the first series all started in $E^u(A)$; the initial states were arbitrarily chosen using a random number generator, with the restriction that the distance of each initial state to A was fixed: if ψ^i denotes an initial state, the restriction was $||\psi^i - A|| = c\,||A - B||$, with $||.|| = <.,. >_1^{1/2}$, the square root of the kinetic energy inner product, and $c = 0.15$. This moderately small choice for c ensures that the model doesn't stay in the linear regime of A for too long, thereby keeping the total computation time within reasonable limits. During each integration the distance to B was monitored and the minimum distance to B was noted. These minimum distances showed large variances: using again the square root of the kinetic energy distance between A and B as unit, the minimum distances varied between 1 and 0.59. In the panels on the left of figure 2.7 the eight orbits that came closest to B (minimum distances between 0.59 and 0.67) are plotted in several projections onto the leading EOFs. They clearly show a preference for a specific region of phase space.

The 500 integrations of the second series were started from arbitrarily chosen initial states in $E^u(B)$ at fixed distance to steady state B. The distances to A were monitored and the minimum distances to A varied between 1 and 0.43. The eight orbits coming closest to A (minimum distances between 0.43 and 0.5) are plotted in the right panels of figure 2.7. Again, they follow roughly similar trajectories. Comparing the left and right panels of figure 2.7 shows marked differences between orbits going from A to B on the one hand and those from B to A on the other hand. In particular PC3 shows a separation: it is mainly negative during the transition from A to B and positive when switching back from B to A.

Although the projections of figure 2.7 suggest that the plotted orbits end up quite close to one fixed point or the other, the numbers show otherwise: the end distances are still of the order of half the distance between A and B. In order to find orbits that end up close to the desired fixed point, we have developed a new algorithm which makes uses of adjoint techniques, stemming from 4D-VAR data assimilation in weather forecasting.

The algorithm is an iterative method. From a given orbit a correction step for the starting point of the orbit is calculated. Applying this correction to the initial state should result in a new orbit ending up closer to the fixed point than the previous one. Suppose we want to find an orbit starting close to A and ending up close to B, and suppose we have a "first guess" orbit $\bar{\psi}(t)$ with $t \in [0, T]$. Let

$$\psi'_f = \mathrm{B} - \bar{\psi}(T) \tag{2.5}$$

and

$$G = <\psi'_f, \psi'_f>_1, \tag{2.6}$$

with $<.,. >_1$ the kinetic energy inner product. The smaller G is, the closer the orbit ends near B, so we want to minimize G. Writing the model equation (2.1) in terms of streamfunction ψ and linearising it around states from the orbit $\bar{\psi}(t)$ yields the tangent linear model, which describes the linear evolution of perturbations ψ^p of $\bar{\psi}(t)$:

$$\dot{\psi^p} = L(\bar{\psi}(t))\,\psi^p \tag{2.7}$$

A second orbit $\tilde{\psi}(t)$ of the full nonlinear system (2.1), starting in $\tilde{\psi}(0) = \bar{\psi}(0) + \psi'(0)$, will end in $\tilde{\psi}(T) = \bar{\psi}(T) + \psi'(T)$. If $\psi'(0)$ is small enough, $\psi'(T)$ will approximately be the result of the tangent linear model (2.7) integrated over time T with initial state $\psi'(0)$.

The formal solution of the tangent linear model can be written using the resolvent $R(t_1, t_2)$ and reads

$$\psi^p(t_2) = R(t_1, t_2)\,\psi^p(t_1) \tag{2.8}$$

If we consider ψ'_f to be a perturbation of the orbit $\bar{\psi}(t)$ at its end-point we can relate it to a perturbation ψ'_i at its starting point by substituting $\psi^p(t_1) = \psi'_i$ and $\psi^p(t_2) = \psi'_f$ in the above equation, giving $\psi'_f = R(0, T)\,\psi'_i$. Writing shortly R for $R(0, T)$ and substituting in equation (2.6) yields

$$G = <\psi'_f, R\,\psi'_i>_1 = <R^*\,\psi'_f, \psi'_i>_1 = -<\Delta R^*\,\psi'_f, \psi'_i>_0 \tag{2.9}$$

in which R^* is the adjoint of R with respect to $< .,. >_1$ (note that the adjoint depends on the inner product) and $< .,. >_0$ is the squared norm inner product (see appendix 2.A). For more information on the adjoint, in particular on the adjoint of the tangent linear barotropic model with respect to the kinetic energy inner product, see Talagrand and Courtier (1987) and Barkmeijer (1992).

We can now relate a change in the initial state of the orbit $\bar{\psi}$ to a change in the distance between the end point of $\bar{\psi}$ and the fixed point B:

$$\frac{\delta G}{\delta \psi'_i} = -\Delta R^*\,\psi'_f \tag{2.10}$$

This gives us the recipe to improve on the orbit $\bar{\psi}$: calculate the difference vector ψ'_f, integrate this vector using the adjoint of the tangent linear model and let the Laplace
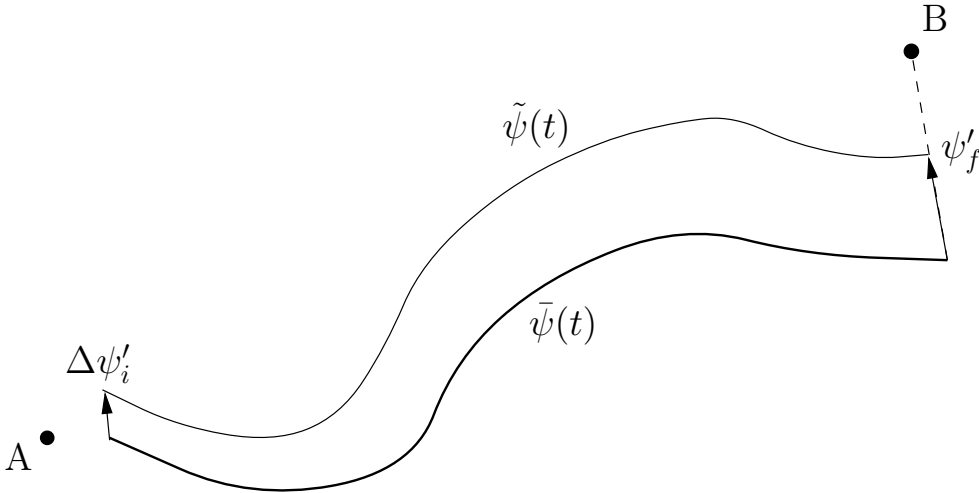
Figure 2.8: *Schematic representation of the adjoint method to find approximate heteroclinic connections. The vector ψ'_f pointing from the end point of the orbit $\bar{\psi}(t)$ to the fixed point B is integrated with the adjoint of the tangent linear model, resulting in ψ'_i. By moving the initial state in the direction of $\Delta \psi'_i$ a new orbit (of the full nonlinear system) $\tilde{\psi}(t)$ is obtained ending closer to B.*

operator Δ act on the resulting vector $R^* \psi'_f$. Shift the initial state $\bar{\psi}(0)$ in the direction of $-\Delta R^* \psi'_f$. The orbit $\tilde{\psi}$, that results from integrating the nonlinear system (2.1) with this new initial condition, should end up closer to the desired fixed point. See figure 2.8 for a schematic representation of the procedure. The method is applied iteratively under the constraint that the initial point of the orbit at each iteration step stays both in the linear approximation of the unstable manifold and at a fixed, small distance to the steady state.

The time-length T of the orbit $\bar{\psi}$ hasn't been considered yet. As was mentioned previously, a genuine heteroclinic orbit stretches over an infinite amount of time; however, the approximation of the connection, from $E^u(A)$ to $E^s(B)$, takes a finite amount of time. This time-length is not known a priori, so the algorithm just described, which tries to find the approximation, must allow for the possibility to adapt the interval T. This is taken care of by starting with a relatively small T and extending the length if at the previous iteration step the algorithm could only marginally improve on the orbit (that is, when the relative decrease of $G^{1/2}$ drops below, say, 1%).

To initialize the algorithm the 2 sets of eight "optimal" initial states, described previously, were taken from the 2 sets of 500 integrations. The "initial guess orbits" were set to a length of 25 days, starting from the two times eight initial states. The magnitude of the correction steps made to the initial states was initially set to a length of 0.1 times the distance between the initial point and the nearby steady state; this magnitude was decreased when lengthening the time interval T.

From both sets we present the orbits showing the best results (that is, the orbits coming closest to the fixed points while the algorithm couldn't improve on them anymore): 2 of the A to B (zonal to blocked regime) orbits have end distances to fixed point B of 0.35 (length: 167 days) and 0.37 (183 days). Compared to the earlier results with minimum distances to B between 0.59 and 0.67, this is a considerable improvement. It must be
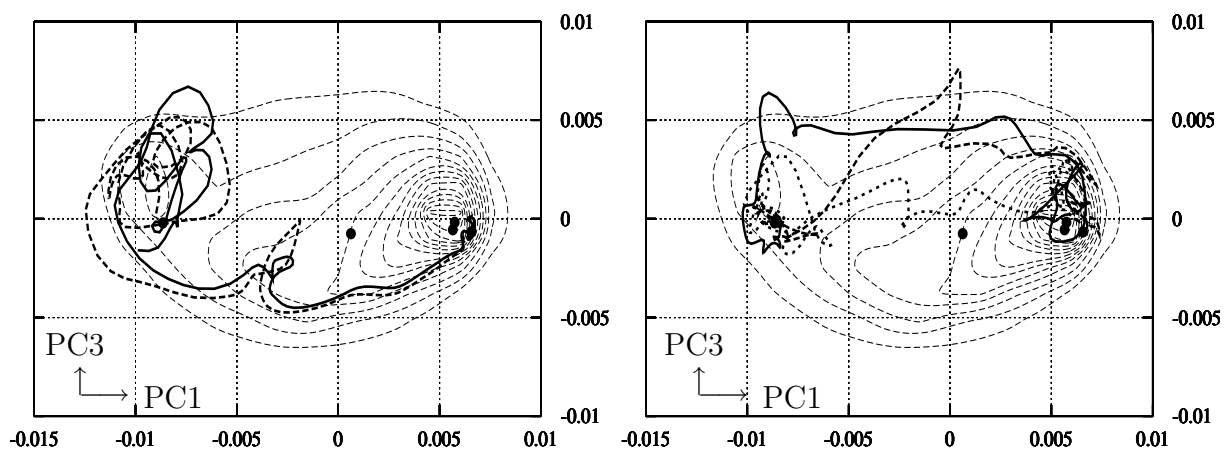

Figure 2.9: *Approximations of nearby heteroclinic connections between regimes. Left: zonal to blocked, right: blocked to zonal. Shown is the projection of the orbits onto the PC1,PC3-plane.*

mentioned here that in the 200 year dataset there are only 9 events in which the distance to B is less than or equal to 0.37; the absolute minimum in the whole dataset is 0.30. For the reverse case (B to A) we have 3 orbits, with end distances to A of 0.15 (173 days), 0.13 (121 days) and 0.11 (217 days). The 200 year dataset has an absolute minimum of 0.11 and 2 events below 0.15. It must be stressed that all 5 orbits are solutions of the full nonlinear model (2.1), integrated with the same integration routine as was used to create the 200 year dataset.

The projections of the 5 orbits onto the PC1,PC3 plane are shown in figure 2.9. The characteristics already apparent in figure 2.7 show up again for these 5 orbits. The numbers show that the B to A orbits come much closer to fixed point A than the A to B orbits come to B. Moreover, the latter two are very similar to each other whereas the former three have clearly distinct trajectories (see figure 2.9). This supports the argument, given earlier on in this section, that the B to A connection is probably more generic (and could even be a 2-dim surface rather than a single curve) than the reverse connection.

2.5.3   Data interpretation

The dynamical structure that emerges by accepting the hypothesis that regime behaviour is linked to heteroclinic dynamics has its consequences for the behaviour of the barotropic model. We have seen that the approximations of the (nearby) heteroclinic connections travel through certain regions of phase space during transitions from one regime to the other. If the full barotropic system, as described by equation (2.1), really is influenced by these heteroclinic connections, this should be visible in the data as well. More specifically, the system should show in its data a tendency to have a negative PC3 anomaly when going from the zonal regime MAX1 to the blocked regime MAX2, and a positive PC3 anomaly when returning from the blocked to the zonal regime, since this is how the heteroclinic connections travel. Furthermore, during transitions from MAX1 to MAX2 the found orbits stay closer together than during transitions in the other direction. One would therefore expect the data of the MAX2 to MAX1 transitions to be more scattered than the data for the reverse transitions. Finally, the region with strongly positive PC2 and weakly negative PC1 anomalies is visited during MAX1 to MAX2 transitions but not during MAX2 to MAX1 transitions.

These assumptions can be checked using the 200 year dataset discussed earlier. Let us define the two regimes in a very coarse way, thereby including probably too many states in these regimes: all states with PC1 > 0.004 count as states in the zonal regime, all states with PC1 < -0.008 are said to be in the blocked regime. All transitions, i.e. all the segments of the data for which -0.008 < PC1 < 0.004, are divided in four categories: transitions from the zonal to the blocked regime (z → b), from blocked to zonal (b → z), from zonal back to itself (z → z) and from blocked back to itself (b → b). The beginning and ends of the transition segments are determined by PC1 crossing the value of 0.004 or -0.008. The duration of the transitions is not taken into account here. From the four different datasets that result PDFs are made. The number of datapoints ending up in each dataset is as follows: z → b: 10548; b → z: 6648; z → z: 26398; b → b: 2894. The projections of the PDFs of the b → z and z → b segments onto the planes of the leading three EOFs is shown in figure 2.10. The EOFs that are used are those of the full 200yr

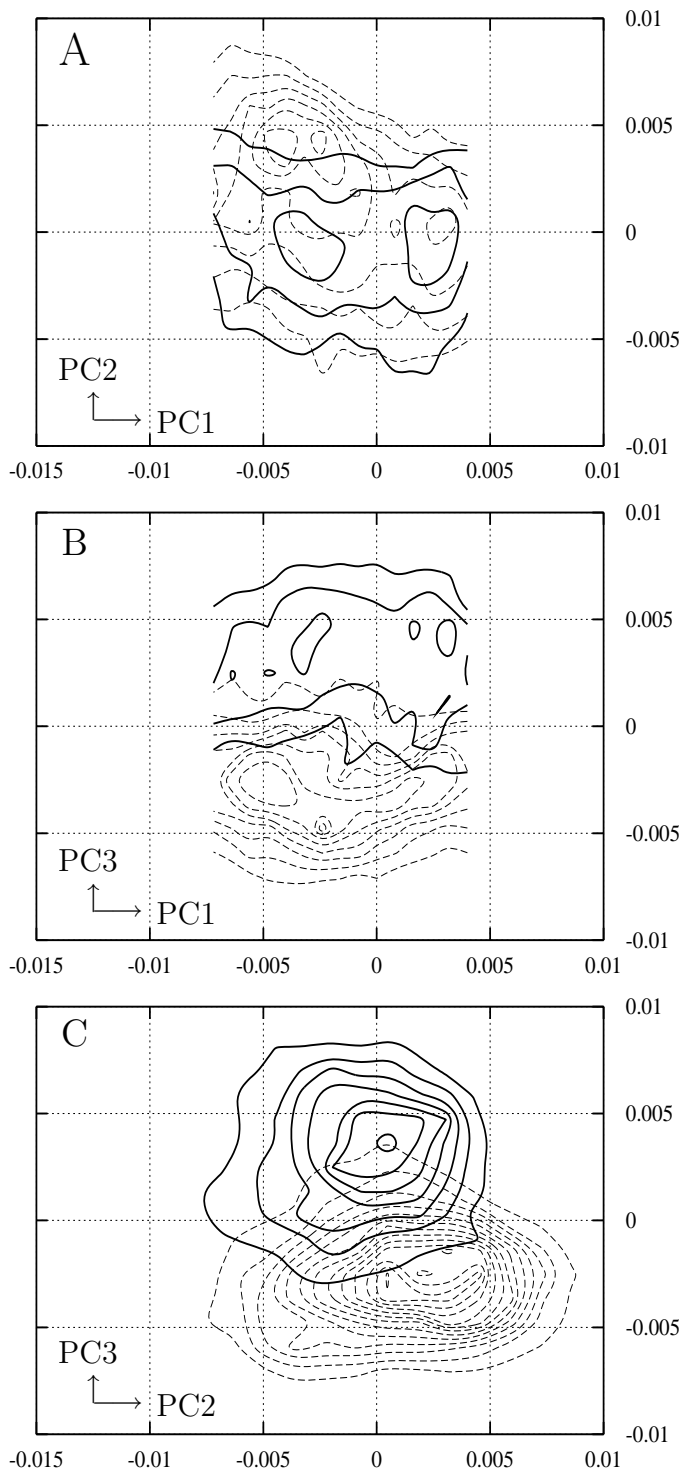Figure 2.10: *PDFs of transitional states in 200yr dataset. Dashed lines: zonal → blocked; solid thick lines: blocked → zonal. Contour interval is 25. Projections shown are PC1 vs. PC2 (A), PC1 vs. PC3 (B) and PC2 vs. PC3 (C)*

dataset. The PDFs show that the data indeed possesses the transition-characteristics that could be predicted from the trajectories of the near-heteroclinic orbits in figures 2.7 and 2.9. PC3 is negative during z → b and positive during b → z transitions. The region with weakly negative PC1, strongly postive PC2 is often visited during z → b but hardly during b → z transitions. The PDFs of the z → b transitions are more confined than those of the b → z transitions (more specifically: the maxima of the z → b PDFs are much more pronounced, showing that these transitions are more confined than the b → z ones).

The b → b and z → z transitions have not been discussed yet. The former has not enough realisations to be really interesting, but the latter does have many realisations. Although these transitions don't involve transitions from one regime to another but rather from the zonal regime back to itself, the concept of global bifurcations can still shed some light on these transitions. If the hypothesis is correct that the z → b → z heteroclinic cycle gets broken and results in a homoclinic z → z connection, one would still expect this homoclinic connection, and hence the data from the z → z transitions, to show traces of the original cycle. The preference for certain phase space regions during the first half (leaving the zonal regime) and second half (returning to the zonal regime) of the z → z transition may still be somewhat similar to the preferences of the z → b resp. the b → z transition. In order to check this, all segments of the z → z transitions were cut in two. All datapoints of the first half of each z → z segment were collected in one dataset, the points of the second halves in another. The PDFs of these two datasets are shown in figure 2.11. The projection on the PC1,PC2-plane doesn't show the characteristics anymore, but the other two projections do: the PC3 anomalies still show the characteristics described earlier: negative when leaving the zonal regime, positive when returning.

The very coarse way in which the two regimes were defined above, and in which the transitions were counted, has the consequence that the "heteroclinic characteristics" that are found back in the data are not very detailed. We could improve on this by selecting the data more precisely, e.g. by defining the regimes more strictly and by imposing a maximum time limit in which the transition from one regime to the other should be completed. However, such stricter selection would obscure the very general influence of the heteroclinic behaviour. The coarse definition and selection of regime- and transition states shows that the heteroclinic characteristics are visible not only for those transitions that run precisely between narrowly defined regimes, but also for a much broader category of states, namely the large majority of transitions between only crudely defined regimes. From the 46488 states that were counted as transitional states, 43549 (93.7%) fell in either the z → b, b → z or z → z cases, which were shown to possess on average the characteristics of heteroclinic transitions.

2.6 Discussion

Looking at the regime behaviour, displayed by the barotropic model, from a dynamical systems point of view gave a new perspective on the dynamics of regime transitions. The thought that these transitions should be related to heteroclinic connections, since the model is entirely deterministic, guided us to the notion that the transitions must be
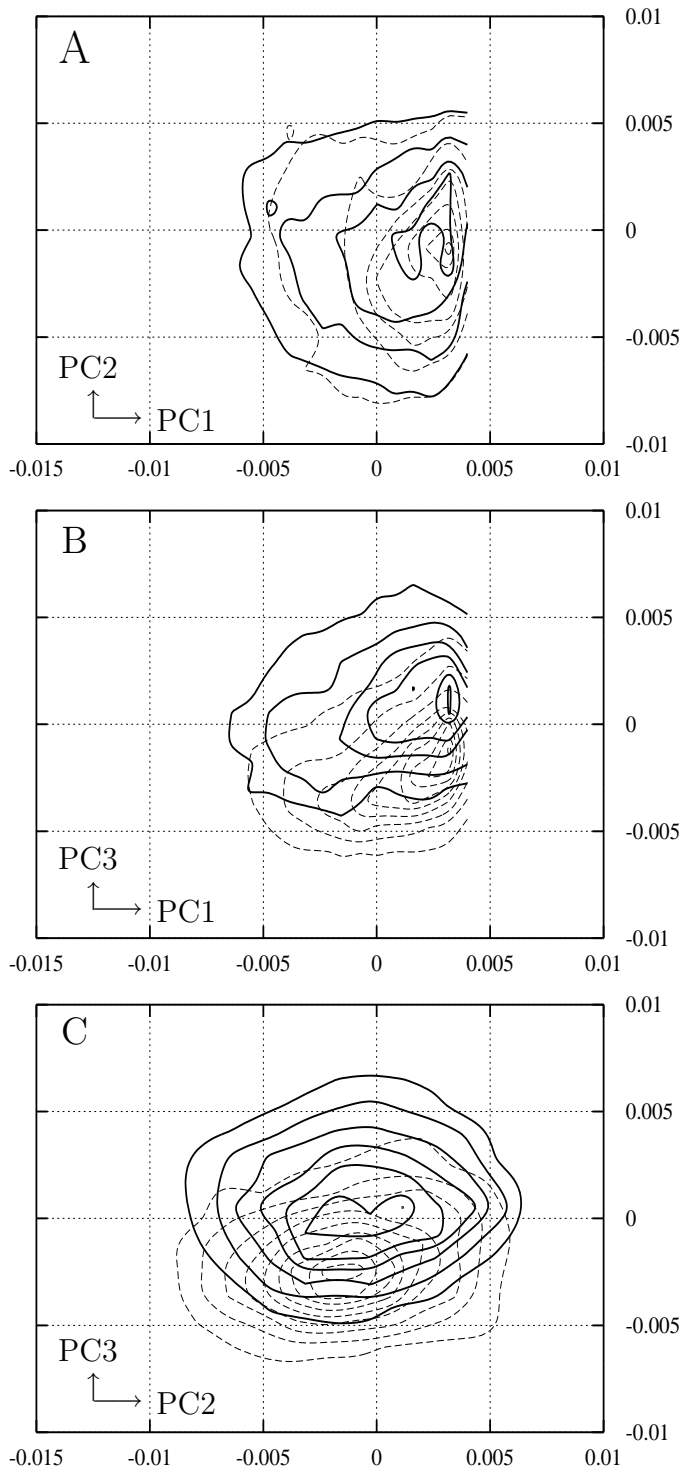
Figure 2.11: *PDFs of zonal → zonal transitional states in 200yr dataset. Dashed lines: first half of each segment; solid thick lines: second half of each segment. Contour interval is 50. Projections shown are PC1 vs. PC2 (A), PC1 vs. PC3 (B) and PC2 vs. PC3 (C)*

structured in phase space. The transitions from the zonal to the blocked regime will, on average, follow paths through different regions of phase space than the transitions from the blocked to the zonal regime. The 200 yr model dataset indeed shows clear differences between the two transitions, even with a coarse filtering; these differences match well with the approximations of (the ruins of) the hypothesized heteroclinic connections. A few remarks must be made to conclude this paper:

• The new algorithm, presented in this study, to approximate the ruins of the heteroclinic connections was able to improve on the optimal transition orbits that resulted from choosing many arbitrary initial states in (the linear approximation of) the unstable manifolds of the two steady states. Nevertheless, the algorithm still needs improvement in order to make it work more easily and probably with better results, in different model contexts. In particular, the handling of the time-interval and the choice of the length of the correction step can be more sophisticated. However, such improvement is beyond the scope of this paper.

• The difference between the two regimes mainly projects onto EOF1, which resembles the Arctic Oscillation pattern. The difference between the z→b and b→z transitions is most clearly visible in EOF3, which projects strongly onto the North Atlantic Oscillation pattern (with positive PC3 corresponding to the negative NAO-phase and vice versa). Combining these findings in one picture results in figure 2.12, a highly simplified, schematic picture of the dynamics of regime behaviour, NAO and AO in the barotropic model. Starting in the zonal regime, which shows a polar vortex that is weaker than
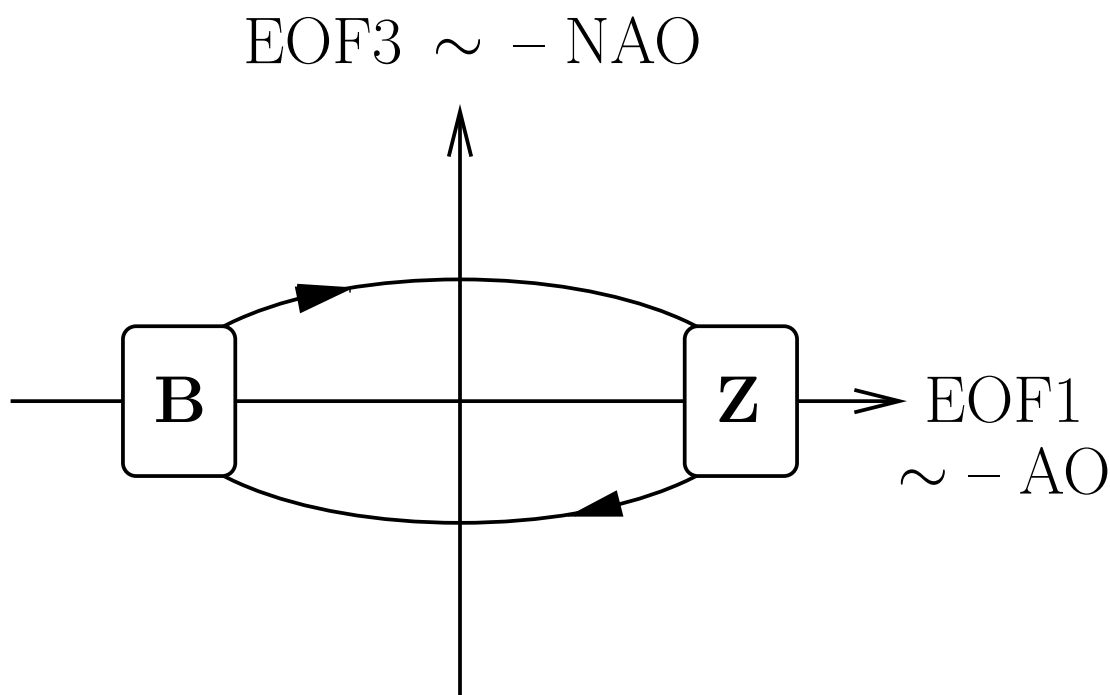

Figure 2.12: *Schematic representation of the interrelationships of regime behaviour, NAO and AO, as emerging from the barotropic model. B denotes the blocked regime, Z the zonal regime. Negative PC1 corresponds to a stronger polar vortex (positive AO phase), negative PC3 corresponds to the positive NAO phase.*

average, the model evolves via a positive NAO phase to the blocked regime. The blocked regime is characterised by a blocking over Europe and, on a larger scale, a strong polar vortex. Leaving the blocked regime, the block persists for a while but shifts to the west, to the North-Atlantic. This transition phase thereby gets the features of a negative NAO phase. Finally, the model atmosphere is back in the zonal regime.

• The great (phase space) variety of the b → z transition orbits compared to their z → b counterparts (which showed up both in the data and in the approximations to the heteroclinic connections, and which can be explained partly by looking at the dimensions of the stable and unstable manifolds of the involved fixed points) raises interesting questions about the formation and destruction of blockings in our model. There seems to be less freedom in the ways blockings can be created than there is in the ways blockings are destroyed. The breaking up of the zonal regime should follow a rather specific scenario for the system to evolve to a blocking situation within a certain amount of time (say two months).

• This conclusion cannot end without pointing out that the model used in this study, a T21 barotropic model with realistic forcing, obviously has its limits: the spatial resolution is rather coarse, it is limited to northern hemisphere flow, and baroclinic dynamics is absent (but note the implicit presence of the time-averaged impact of neglected (e.g. baroclinic) processes in the forcing). Nevertheless, the bimodality and regimes observed in this model have several realistic aspects. Blocked flows, variations of the polar vortex and the NAO are all known features of the atmospheric flow. The details may be inaccurate but in essence these phenomena of the barotropic model are realistic. The regimes MAX1 and MAX2 resemble the AO in its negative and positive phase, respectively (Thompson and Wallace, 1998). They combine a weakening resp. strengthening of the wintertime polar vortex with anomalies of the same sign (but opposite to the polar anomaly) over Europe and the North-Pacific (see figure 2.4). In the real atmosphere, the anomaly over Europe stretches towards North-America. This stretching is absent in the regimes of the barotropic model, but somewhat present in our EOF1. Our regimes can also be compared to the clusters found by Corti et. al. (1999). Their clusters D and B resemble our MAX1 and MAX2. Furthermore, the pattern of their cluster A is comparable to a negative PC3 anomaly of our model: positive NAO-phase, negative anomaly over the North-Pacific, positive anomaly over North-America. These similarities beg the question whether transitions from cluster D to B have a tendency to go via cluster A, which would be the same order as we found in the barotropic model. In our model data no regime was found that corresponds to cluster A. However, the preferred transition routes have a small "loop" at the midpoint of the MAX1 to MAX2 transition (figure 2.9). This extra curvature may result in a local PDF maximum in time-averaged data (note that Corti et. al. use monthly mean data, whereas we have used daily data). The PDFs for the z → b transitional states (figure 2.10) have indeed a maximum at the location of the small loop. This observation suggests that there may be two types of regimes: one type associated with (marginally unstable) equilibria of the large-scale flow, the other type induced by local curvatures of the preferred transition orbits. Finally, the resemblance of cluster C and our positive PC3 anomaly is less strong (although not completely absent), so the return transition from B to D via C may be less clear. This could be related to our findings that in the barotropic model the MAX2 to MAX1 transitions are less confined than the reverse ones.

We have argued that the regime behaviour of the barotropic model is quite realistic in its qualitative aspects. The concept of heteroclinic connections gave a new perspective on the regime behaviour in the model. It was shown that these connections have important implications for the details of the regime transitions, an aspect which has been largely ignored in the literature. This perspective can very well be used for further, more detailed studies of transitions in both models and observations.

2.A   Appendix: The inner product

The inner product used throughout this paper is the kinetic energy inner product, defined by

$$< \psi_1, \psi_2 >_1 = \frac{1}{2} \int (\nabla \psi_1) \cdot (\nabla \psi_2) \, d\Omega \qquad (2.A.1)$$

The integral is taken over the entire sphere; $d\Omega$ denotes a volume element on the sphere. ψ_1 and ψ_2 are streamfunction fields on the sphere. For the barotropic model, the kinetic energy of a state with streamfunction ψ is given by $< \psi, \psi >_1$. The kinetic energy inner product belongs to a class of inner products

$$< \psi_1, \psi_2 >_k = \frac{1}{2} \int (\nabla^k \psi_1) \cdot (\nabla^k \psi_2) \, d\Omega, \quad k \in \mathbb{N} \qquad (2.A.2)$$

It is easy to show that $< \psi_1, \psi_2 >_1 = - < \xi_1, \psi_2 >_0$, where $\xi_1 = \Delta \psi_1$, the vorticity field associated with ψ_1.

2.B   Appendix: Gradient and Jacobian

The scalar function F to be minimized in section 2.4 is the discretized version of the functional $\mathcal{F}$, which is defined as the squared norm of the vorticity tendency integrated over the sphere:

$$\mathcal{F} = \int (\dot{\xi})^2 \, d\Omega. \qquad (2.B.1)$$

All integrals in this appendix are assumed to be taken over the sphere; $d\Omega$ denotes again a volume element on the sphere. Using the definition $\delta \mathcal{F}(\xi) = \mathcal{F}(\xi + \delta \xi) - \mathcal{F}(\xi)$ and

substituting equation (2.1) gives:

$$
\begin{aligned}
\delta\mathcal{F} &= 2\int \dot{\xi}\,(\delta\dot{\xi})\,\mathrm{d}\Omega \\
&= 2\int \dot{\xi}\,(J(\delta\xi,\psi) - J(\delta\psi,\xi+f+h) + k_1\Delta\delta\psi + k_2\Delta^4\delta\psi)\,\mathrm{d}\Omega \\
&= 2\int \delta\psi\,(J(\dot{\xi},\xi+f+h) + \Delta J(\psi,\dot{\xi}) + k_1\Delta\dot{\xi} + k_2\Delta^4\dot{\xi})\,\mathrm{d}\Omega \quad\quad (2.\mathrm{B}.2)
\end{aligned}
$$

The above derivation uses $\delta\xi = \Delta\delta\psi$, as well as the following identities:

$$
\int \eta\,(\Delta\zeta)\,\mathrm{d}\Omega = \int \zeta\,(\Delta\eta)\,\mathrm{d}\Omega \quad\quad (2.\mathrm{B}.3)
$$

$$
\int \eta\,J(\zeta,\nu)\,\mathrm{d}\Omega = -\int \nu\,J(\zeta,\eta)\,\mathrm{d}\Omega \qu\quad\quad (2.\mathrm{B}.4)
$$

These identities stem from the identity $\int \nabla\cdot\mathbf{A}\,\mathrm{d}\Omega = 0$, which is valid for any smooth vector field $\mathbf{A}$ since the integration runs over the sphere.

From equation (2.B.2) follows

$$
\frac{\delta\mathcal{F}}{\delta\psi} = 2\left(J(\dot{\xi},\xi+f+h) + \Delta J(\psi,\dot{\xi}) + k_1\Delta\dot{\xi} + k_2\Delta^4\dot{\xi}\right) \qu\quad\quad (2.\mathrm{B}.5)
$$

Discretizing the equations, i.e. expressing all fields in terms of a finite number of spherical harmonics, yields the approximation F of $\mathcal{F}$; ψ is replaced by $(\psi_1,\ldots,\psi_N)^{\mathrm{T}}$. The functional derivative (2.B.5) becomes the vector $\partial F/\partial\psi_i$, $i=1,\ldots,N$: the gradient of F with respect to $(\psi_1,\ldots,\psi_N)^{\mathrm{T}}$. This is the gradient needed for the minimization algorithm.

An expression for the Jacobian, the matrix that describes the model linearised around some state, can be derived in a similar way as the expression of the gradient above. The Jacobian (denoted by L) is defined as

$$
L_{ij} = \frac{\partial\dot{\psi}_i}{\partial\psi_j} \qu\quad\quad (2.\mathrm{B}.6)
$$

which may be thought of as a series of gradients of scalars $\dot{\psi}_i$. Define

$$
D_i = \dot{\psi}_i = \int Y_i^*\,\dot{\psi}\,\mathrm{d}\Omega \qu\quad\quad (2.\mathrm{B}.7)
$$

in which Y_i^* is the complex conjugate of the spherical harmonic with (multi-)index i; the gradient can be shown to be

$$
\frac{\delta D_i}{\delta\psi} = J(\Delta^{-1}Y_i^*,\xi+f+h) + \Delta J(\psi,\Delta^{-1}Y_i^*) + k_1 Y_i^* + k_2\Delta^3 Y_i^* \qu\quad\quad (2.\mathrm{B}.8)
$$

Discretizing in spherical harmonics gives the Jacobian:

$$
L_{ij} = \left(\frac{\delta D_i}{\delta\psi}\right)_j \qu\quad\quad (2.\mathrm{B}.9)
$$

Chapter 3

A mechanism for atmospheric regime behaviour

Abstract. Adopting the viewpoint that atmospheric flow regimes can be associated with steady states, we investigate the hypothesis that regime transitions in deterministic atmosphere models are related to the existence of heteroclinic connections between these steady states. We study a low-order barotropic model with topography, in which topographic and barotropic instabilities are the mechanisms dominating the dynamics. By parameter tuning, the Hopf bifurcation corresponding to barotropic instability can be made to coincide with one of the saddle-node bifurcations that are due to the topography in the model. This coincidence is called a fold-Hopf bifurcation. Among the dynamical structures related to such a bifurcation are heteroclinic connections and homoclinic orbits, connected to the equilibria. A heteroclinic cycle back and forth between the equilibria, present in the unfolding of the truncated normal form of the fold-Hopf bifuraction, will be perturbed in the full model, leaving orbits homoclinic to one of the equilibria. The impact of these mathematical structures explains several characteristics of regime behaviour known from previous model studies.

3.1 Introduction

More than fifty years after the first reports on the topic, the regime behaviour of the atmosphere remains an enigma. The considerable attention that has been devoted to the hypothesis that atmospheric low-frequency variability is affected by the existence of preferred flow regimes has not yet resulted in a fully conclusive picture. Nevertheless, the detection of regimes in the observational data of the atmosphere has progressed a lot in the last two decades. Starting with Benzi et al. (1986) and Sutera (1986), the somewhat intuitive notion of regimes, developed since the works of Namias (1950) and Rex (1950a,b), was put on a firmer basis using concepts and techniques from modern probability theory. In papers such as Cheng and Wallace (1993), Kimoto and Ghil (1993a,b), Smyth et al. (1999) and Monahan et al. (2001), the use of these techniques has resulted in growing evidence for the existence of regimes in northern hemispheric atmosphere data. Moreover, the regimes found in these papers, using different techniques, are strikingly similar.

Accepting the existence of regimes still leaves us with the question which dynamical processes are responsible for this behaviour. An important contribution regarding this

issue was made by Charney and DeVore (1979, CdV from now on), who stated that flow regimes should be identified with equilibrium solutions of the equations describing the evolution of large scale atmospheric flow. Their hypothesis, supported by the study of a low-order model for barotropic flow over topography, was taken up and expanded on by, among others, Reinhold and Pierrehumbert (1982), Legras and Ghil (1985), Källén (1981, 1982), Yoden (1985), De Swart (1988a,b, 1989) and Itoh and Kimoto (1996, 1999).

Although the above model studies unveiled much about the structure of the (model) regimes, the observed regime transitions have so far not been satisfactorily explained. From statistical studies of model data as well as observational data, it has been known for some time that transitions between regimes are not an entirely random process. In Mo and Ghil (1987, 1988), Molteni et al. (1990) and Kimoto and Ghil (1993b), up to seven regimes/clusters are identified in various datasets (both from observations and models). By counting the transitions between the various regimes, these studies made clear that there exist not only preferred flow regimes, but also preferred transitions between (some of) the regimes. Plaut and Vautard (1994), studying the interplay between low-frequency oscillations and regimes using MSSA, also found that transitions are not purely random.

Itoh and Kimoto (1996, 1997, 1999) propose chaotic itinerancy as an explanation for the preferences and inhomogeneities in regime transitions. Using multi-layer quasi-geostrophic models of moderate complexity (L2T15 and L5T21), they detect multiple attractors that are identified as regimes. By changing a parameter (horizontal diffusivity or static stability), these attractors lose their stability one by one, thereby admitting regime transitions. When all attractors have lost their stability, the model shows irregular transitions between the remnants of the former attractors, and thus between the regimes. This behaviour is called chaotic itinerancy. The preferred order of the transitions is related to the order in which the attractors lost their stability. This interesting result once more emphasizes the inhomogeneities in regime transitions, which must be due to the nonlinear nature of the system under study. However, it does not provide much insight in the dynamics that determine the transitions, as the loss of stability of the various attractors is not analysed. It remains unknown what dynamical processes cause and drive the transitions.

Often a stochasticity assumption is invoked to explain regime transitions. Noise, representing the effect of unresolved physics and dynamics, can kick a system out of the basin of attraction of one regime and into another. This has been studied, by adding stochastic perturbations to a low-order model, in e.g. Egger (1981), Benzi et al. (1984), De Swart and Grasman (1987) and Sura (2002). However, the addition of noise is not necessary to trigger regime transitions. Even in deterministic low-order models transitions can occur. It is therefore worthwhile to consider how transitions can be generated by the internal, deterministic dynamics of a system, in the absence of noise.

Our hypothesis will be that barotropic flow over topography is not only sufficient to create multiple equilibria (as in CdV), but can also generate transitions between those equilibria, resulting in regime behaviour. This is based on the fact that transitions have been observed in various barotropic models. Examples are the 25-variable model used by Legras and Ghil (1985), the 10-dimensional model used by De Swart (1988a,b), and the T21 model studied in Crommelin (2002b). The presence of baroclinic processes is apparently no *conditio sine qua non* for regime transitions. Recently, even in laboratory

experiments designed to mimic as closely as possible the situation of barotropic flow over topography in the atmosphere, regime behaviour was observed (Weeks et al. 1997, Tian et al. 2001). Thus, the transitions in those truncated barotropic models cannot simply be discarded as model artefacts.

In Crommelin (2002b) strong numerical evidence was found in support of the hypothesis that remnants of heteroclinic connections are responsible for the transitions between flow regimes. The regimes were found to correspond with steady states, in accordance with the paradigm introduced by CdV, and are likely to have deterministic connections running back and forth between them for nearby parameter values. An attempt was made to find the trajectories of the connections, and the result of that attempt was shown to be consistent with the phase space preferences of regime transitions during a 200 year model integration. In spite of the numerical evidence, analytical, or at least more rigorous mathematical support is still lacking for the hypothesis that regime transitions are related to heteroclinic connections (a hypothesis also mentioned by Legras and Ghil (1985), Kimoto and Ghil (1993b) and Weeks et al. (1997)). In this paper we want to provide such support, by studying a low-order model of the atmospheric flow at midlatitudes. We hypothesize that the combination of topographic and barotropic instability is sufficient to create multiple equilibria corresponding to regimes as well as connections between these equilibria, and we therefore take the simplest model possible that combines these two instability mechanisms. Such a model is provided by De Swart (1989, DS from now on), a six-variable model that is essentially the same as the model of CdV, also studied by Yoden (1985), except for a different scaling and a more general zonal forcing profile. We will use this model to study the interaction between topographic and barotropic instability, in order to see if and how the combination of these instability mechanisms can generate connections between steady states, resulting in regime behaviour.

The study of this interaction can provide a first step towards a better understanding of the mechanisms playing a role in the phenomenon of regime transitions. Although the model that will be used is probably too simple to arrive at conclusions that pertain immediately to the real atmosphere, the insights it provides may guide the investigation of more complex models, or even observational data.

3.2 Derivation of the low-order model

The starting point for this study is the hypothesis that the combination of topographic and barotropic instability is sufficient to generate regime behaviour. Notably, baroclinic processes are not considered to be truly necessary for regime transitions to occur. Thus, we consider a model without baroclinic dynamics and without stochastic terms supposed to represent the effect of baroclinic processes. Instead, we wish to study the simplest deterministic model possible that combines the mechanisms of barotropic and topographic instability. Therefore we use the model presented in DS, a slightly different version of the CdV model. It can have multiple equilibria, caused by topography, and is forced by a zonal flow profile that can be barotropically unstable.

The model is obtained by a Galerkin projection and truncation of the barotropic vorticity equation (BVE) on a β-plane channel. The BVE, a partial differential equation,

reads

$$\frac{\partial}{\partial t}\nabla^2\psi = -\mathcal{J}(\psi, \nabla^2\psi + f + \gamma h) - C\nabla^2(\psi - \psi^*). \qquad (3.1)$$

Time (t), longitude (x) and latitude (y) can take on values $(t, x, y) \in \mathbb{R} \times [0, 2\pi] \times [0, \pi b]$. The parameter $b = 2B/L$ determines the ratio between the dimensional zonal length L and meridional width B of the channel. The streamfunction field $\psi(t, x, y)$ is periodic in x: $\psi(t, x, y) = \psi(t, x+2\pi, y)$. The restriction to the beta plane implies that at the meridional boundaries $y = 0, \pi$ both $\partial\psi/\partial x = 0$ and $\int_0^{2\pi}(\partial\psi/\partial y)dx = 0$. The Coriolis parameter f generates the beta-effect in the model. Orography enters with h, the orographic height, and is scaled with γ. The Jacobi operator $\mathcal{J}$ acts on two fields, say $A(x, y)$ and $B(x, y)$, as follows: $\mathcal{J}(A, B) = \frac{\partial A}{\partial x}\frac{\partial B}{\partial y} - \frac{\partial A}{\partial y}\frac{\partial B}{\partial x}$. Finally, the model is driven by a Newtonian relaxation to the streamfunction profile ψ^*, with damping coefficient C.

To arrive at a finite-dimensional model, the BVE is projected onto a set of basis functions which are eigenfunctions of the Laplace-operator ∇^2. On the chosen rectangular domain, with the abovementioned boundary conditions, these functions are double Fourier modes:

$$\begin{aligned}\phi_{0m}(y) &= \sqrt{2}\cos(m\,y/b) \\[2mm] \phi_{nm}(x, y) &= \sqrt{2}\,\mathrm{e}^{inx}\sin(m\,y/b)\end{aligned} \qquad (3.2)$$

in which $|n|, m = 1, 2, \ldots$. The streamfunction and the topographic height are expanded in this basis:

$$\psi(t, x, y) = \sum_{n,m}\psi_{nm}(t)\,\phi_{nm}, \quad h(x, y) = \sum_{n,m}h_{nm}\,\phi_{nm}, \qquad (3.3)$$

The six-dimensional model is obtained by truncating the expansion after $|n| = 1$ and $m = 2$, so the only remaining basis functions are $\phi_{01}, \phi_{02}, \phi_{11}, \phi_{-11}, \phi_{12}$ and ϕ_{-12}. The time-dependent variables $\psi_{01}, \psi_{02}, \psi_{\pm 11}, \psi_{\pm 12}$ are transformed to real variables:

$$\begin{aligned}x_1 &= \frac{1}{b}\psi_{01}, \quad x_2 = \frac{1}{b\sqrt{2}}(\psi_{11} + \psi_{-11}), \quad x_3 = \frac{\mathrm{i}}{b\sqrt{2}}(\psi_{11} - \psi_{-11}), \\[3mm] x_4 &= \frac{1}{b}\psi_{02}, \quad x_5 = \frac{1}{b\sqrt{2}}(\psi_{12} + \psi_{-12}), \quad x_6 = \frac{\mathrm{i}}{b\sqrt{2}}(\psi_{12} - \psi_{-12}).\end{aligned} \qquad (3.4)$$

The topography h is chosen to have a (1,1) wave profile:

$$h(x, y) = \cos(x)\,\sin(y/b) \qquad (3.5)$$

so the only nonzero topographic expansion coefficients are $h_{11} = h_{-11} = 1/(2\sqrt{2})$. The forcing profile ψ^*, finally, is purely zonal, i.e. $\psi^* = \psi^*(y)$. For the 6D model this means that the only forcing terms will be x_1^* and x_4^*.

The set of ordinary differential equations describing the temporal evolution of the x_i is of the form $\dot{\mathbf{x}} = \mathbf{F} + A\mathbf{x} + B(\mathbf{x}, \mathbf{x})$, with energy- and enstrophy-preserving quadratic

nonlinearities. In detail it reads:

$$
\begin{aligned}
\dot{x}_1 &= & \gamma_1^* \, x_3 & \quad - C\,(x_1 - x_1^*) & \\
\dot{x}_2 &= -(\alpha_1 \, x_1 - \beta_1)\, x_3 & & \quad - C\, x_2 & - \delta_1 \, x_4 \, x_6 \\
\dot{x}_3 &= (\alpha_1 \, x_1 - \beta_1)\, x_2 & - \gamma_1 \, x_1 & \quad - C\, x_3 & + \delta_1 \, x_4 \, x_5 \\
\dot{x}_4 &= & \gamma_2^* \, x_6 & \quad - C\,(x_4 - x_4^*) & + \varepsilon\,(x_2\, x_6 - x_3\, x_5) \\
\dot{x}_5 &= -(\alpha_2 \, x_1 - \beta_2)\, x_6 & & \quad - C\, x_5 & - \delta_2 \, x_4 \, x_3 \\
\dot{x}_6 &= (\alpha_2 \, x_1 - \beta_2)\, x_5 & - \gamma_2 \, x_4 & \quad - C\, x_6 & + \delta_2 \, x_4 \, x_2
\end{aligned}
\tag{3.6}
$$

The various coefficients in these equations are given by

$$
\begin{aligned}
\alpha_m &= \frac{8\sqrt{2}}{\pi}\frac{m^2}{4m^2-1}\frac{b^2+m^2-1}{b^2+m^2}, & \beta_m &= \frac{\beta b^2}{b^2+m^2}, \\
\delta_m &= \frac{64\sqrt{2}}{15\pi}\frac{b^2-m^2+1}{b^2+m^2}, & \gamma_m^* &= \gamma\,\frac{4m}{4m^2-1}\frac{\sqrt{2}b}{\pi}, \\
\varepsilon &= \frac{16\sqrt{2}}{5\pi}, & \gamma_m &= \gamma\,\frac{4m^3}{4m^2-1}\frac{\sqrt{2}b}{\pi(b^2+m^2)}.
\end{aligned}
\tag{3.7}
$$

In the equations, the terms multiplied by the α_i model the advection of the waves by the zonal flow. The β_i-terms are due to the Coriolis force; the γ-terms are generated by the topography. The C-terms take care of the Newtonian damping to the zonal profile $x^* = (x_1^*, 0, 0, x_4^*, 0, 0)$. The δ- and ε-terms describe the nonlinear triad interaction between the zonal $(0,2)$ mode and the $(1,1)$ and $(1,2)$ waves. This triad is responsible for the possibility of barotropic instability of the $(0,2)$ mode.

The number of free parameters in the model equations is six: the damping coefficient C, the forcing parameters x_1^* and x_4^*, the length-width ratio of the beta-channel, b, the Coriolis parameter β and the scaling of the topographic height γ. In the analysis of this paper we will most of the time use the ratio r between x_4^* and x_1^* as a free parameter instead of x_4^* itself. This is done by putting $x_4^* = r\, x_1^*$.

The model parameters x_1^*, x_4^* (or r) and γ will be varied throughout this study. The remaining parameters β, b and C will be kept fixed. The choice $\beta = 1.25$ corresponds to a channel centered around a latitude of $45°$, see also DS. The channel width-length ratio b will be set to 0.5, which is different from the value taken throughout most of the study of DS. In there, $b = 1.6$ was chosen. However, at that value the model contains pitchfork bifurcations that create additional branches of equilibria. Since we want to isolate the mechanisms of barotropic and topographic instability in their simplest form, such additional bifurcations are unwanted. By choosing $b = 0.5$ (i.e. reducing the north-south extent of the channel from 80% to 25% of its east-west extent) these extra pitchfork bifurcations do not occur. In DS, C was furthermore set to 0.1, corresponding to a damping timescale of 10 days (the nondimensional time in the equations was scaled such that $\Delta t = 1$ corresponds to roughly one day). We will adopt the same value for C.

3.3 Topographic and barotropic instability

Nonzero topography may introduce a mechanism of instability in the system. The interaction between the zonal flow and the topography can generate standing wave solutions.

The (0,1) zonal flow component will excite a wave with the same wavenumber as the topography, in this case the (1,1) wave. The effect of the topography doesn't end there; as can be seen in the equations, the (1,2) wave mode is coupled to the (0,2) zonal mode via the topography. It must be pointed out that the topography does not create an oscillatory mode (as is usually the case with instabilities), but rather a resonant response curve which corresponds to a parameter range with multiple equilibria. These equilibria are associated with three different ways in which the advection of relative vorticity, the advection of planetary vorticity and the vortex stretching caused by flow over topography, can balance. One of the equilibria is unstable. Since the instability of this steady state is entirely due to the topography, and not to other mechanisms such as barotropic instability or wave instability, the effect of the topography here has been given a new name, topographic instability.

The barotropic instability is an instability mechanism in the more common sense of the word, as it refers to a situation where a steady state (a mainly zonal flow) loses its stability while a (stable) oscillatory wave mode (a travelling wave) is created. The zonal flow profile must obey certain conditions for this instability to be possible, see e.g. Kuo (1949) and Cushman-Roisin (1994). In particular, the profile must have at least one inflection point (a change in the sign of the total vorticity) in meridional direction. For that reason it is necessary to have more zonal modes resolved in the model, since the (0,1) zonal mode can never be barotropically unstable by itself (it does not obey the inflection point condition). The (0,2) zonal mode can become unstable, though only when $b^2 < 3$, due to Fjørtofts theorem (see DS). Our choice of $b = 0.5$ clearly satisfies this condition.

The effect of the two mechanisms is shown in figure 3.1. On the left is a curve of fixed points depicted in the situation without either mechanism playing a role. This was achieved by putting $\gamma = x_4^* = 0$, so that topography is zero and the zonal flow profile
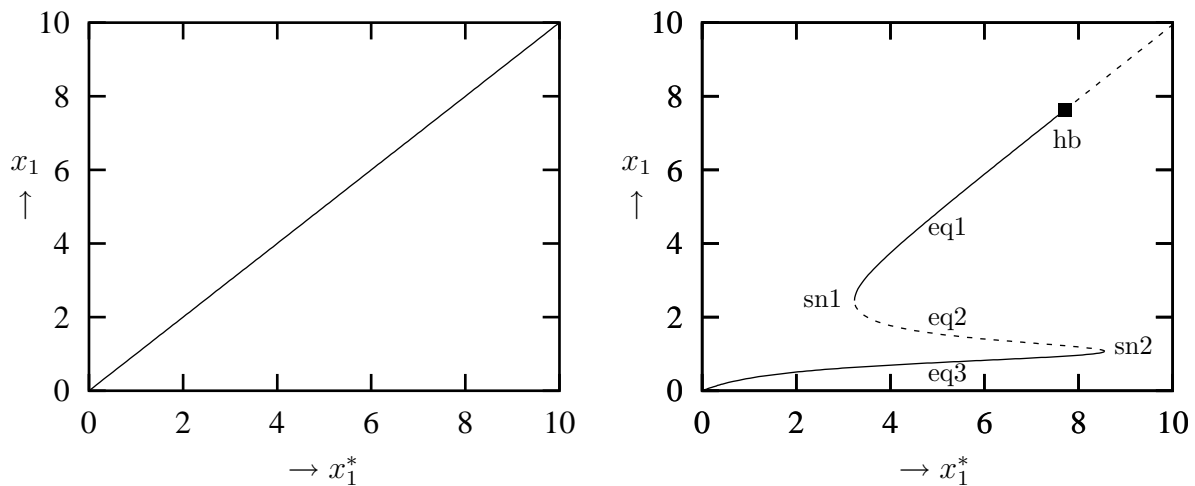
Figure 3.1: *Continuation of fixed points. Solid lines denote stable branches and dashed lines unstable branches. On the left, barotropic instability is not possible and topography is zero ($r = \gamma = 0$). On the right, topography is nonzero ($\gamma = 1$), generating two saddle-node bifurcations (sn1 and sn2). The three branches of equilibria are denoted by eq1, eq2 and eq3. The Hopf bifurcation hb (■) is due to barotropic instability ($r = -0.4$).*

cannot be barotropically unstable. The response of the model to varying x_1^* is a shift of the steady state such that $x_1 = x_1^*$. On the right, both mechanisms are present. The deformation of the straight curve on the left to the S-shaped curve on the right, involving two saddle-node bifurcations (sn1 and sn2), is the result of non-zero topography ($\gamma = 1$). The (supercritical) Hopf-bifurcation on the upper part of the branch represents the barotropic instability, which was triggered by non-zero x_4^*. To be able to perform fixed point continuations starting with zero flow (all $x_i = 0$) at zero forcing, x_4^* is scaled with x_1^*, by putting $x_4^* = r\, x_1^*$ and fixing r at some value (thus, r then controls the shape of the forcing profile while x_1^* determines its strength). This scaling will be used throughout this paper. In the right panel of figure 3.1, $r = -0.4$ was taken. The continuations were carried out using the software package AUTO (Doedel and Wang, 1995).

The flow patterns corresponding to the three different fixed points existing at $x_1^* = 6$ in figure 3.1 are shown in figure 3.2. As was already known from previous studies (e.g. DS), the upper branch eq1 corresponds to equilibria with largely zonal character. The other two are dominated by topographically excited standing wave patterns, one superresonant (eq2), the other subresonant (eq3). The former has a slightly stronger zonal flow component than the latter. The phases of their wave components with respect to the topography are somewhat different: the wave pattern of eq2 is a bit more upstream than that of eq3. The phase difference goes to zero when moving towards sn2.
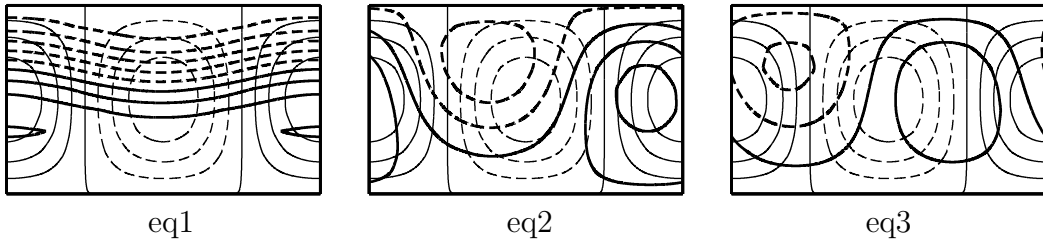

eq1   eq2   eq3

Figure 3.2: *Streamfunction patterns corresponding to the equilibria at $x_1^* = 6, r = -0.4, \gamma = 1$. On the left and the right are the stable equilibria, in the middle the unstable one. Thick lines are streamfunction lines (contour interval 1 in nondimensional units), thin lines are topography contours (interval 0.25 km). Dashed contours are for negative values, solid contours for zero or positive values.*

The subresonant solution eq3 is usually the one that is identified as the regime of blocked flow. The main reason for that seems to be the fact that in the simplest setting (that is, topographic instability being the only physical mechanism present) eq3 is the stable solution, whereas eq2 is unstable. Nevertheless, eq2 also possesses the characteristics of a blocked flow regime. To call eq2 an intermediate solution only because it is unstable in the simple setting seems premature. Eventually, all equilibria have to be unstable for regime transitions to occur in a model without stochastic terms. We see no a priori reason to exclude eq2 as a candidate for the blocked regime.

3.4   The merging of two instabilities: a fold-Hopf bifurcation

3.4.1   Instabilities and bifurcations

The two instabilities are, mathematically, represented by saddle-node- and Hopf-bifurcations. The locations of these bifurcations depend on the parameters of the model. Figure 3.1 already gave an idea of this dependence. The straight curve of fixed points, where the equilibrium is such that the dissipation and the forcing of the first zonal mode are in balance, gets deformed by increasing the amplitude of the topography above some threshold value. At the threshold value, the two saddle-node bifurcations sn1 and sn2 coincide. Thus, the threshold is a cusp-bifurcation. The cusp is a codimension two bifurcation (two parameters, say x_1^* and γ, must be tuned for this bifurcation to occur) and can be continued adding a third parameter, say r. This means that the value of γ for which the cusp occurs changes when altering the shape of the zonal forcing. A curve showing the γ-location of the cusp for varying r is drawn in figure 3.3. It must be stressed that the value of x_1^* for which the cusp occurs also changes along the curve: it decreases monotonically from $x_1^* = 1.992$ at $\gamma = 0.3198$ to $x_1^* = 0.3320$ at $\gamma = 3.0 \times 10^{-12}$. The continuation of the cusp was performed using the continuation software package CONTENT (Kuznetsov and Levitin, 1997). Note that if the zonal forcing profile has the right shape and strength ($x_1^* = 0.3320, r = -0.861467$), the influence of the topography is even felt when its amplitude is infinitesimally small (γ approaching zero).
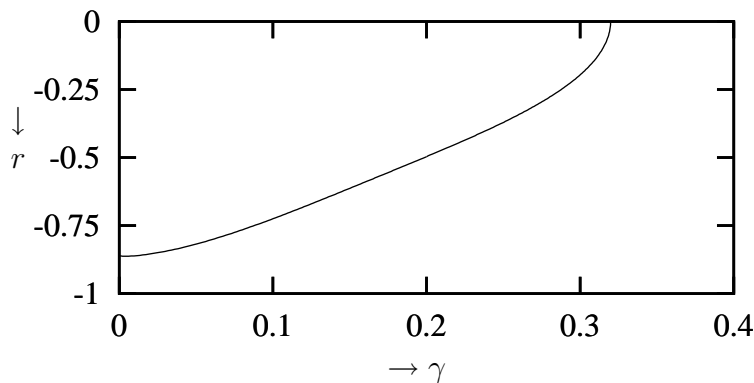

Figure 3.3: *Location of cusp bifurcation for varying γ and r. The value of x_1^* also changes along the curve but is not shown.*

The Hopf-bifurcation is the dynamical expression of barotropic instability. Since the $(0,2)$ zonal mode can become barotropically unstable but the $(0,1)$ mode cannot, it can be expected that the bifurcation occurs at smaller x_1^* values if $|r|$ is increased. By doing so, the Hopf bifurcation will approach the saddle-node sn1, and at some point coincide with it. This simultaneous occurrence of a saddle-node- (or fold) and a Hopf bifurcation is called a fold-Hopf bifurcation, and has codimension two.

Let us investigate what happens when this fold-Hopf bifurcation occurs (i.e. when the two instability mechanisms merge). The description of the various dynamical phemonena

and structures that emerge out of this codimension two point in parameter space, the so-called unfolding of the bifurcation, is given in Kuznetsov (1995). We will briefly review it here.

3.4.2   The fold-Hopf bifurcation: some theory reviewed

The occurrence of a bifurcation of a fixed point can be read off from the eigenvalue-spectrum of the fixed point. In a saddle-node- (or fold-) bifurcation, one eigenvalue is exactly zero, all the others have non-zero real parts. In a Hopf-bifurcation, one complex conjugated pair of eigenvalues has real part zero (and imaginary part non-zero); again, all other eigenvalues are bounded away from the imaginary axis. It will come as no surprise that a fold-Hopf bifurcation is characterised by a fixed point eigenvalue spectrum with one zero and one purely imaginary pair (i.e. $\lambda_1 = 0, \lambda_{2,3} = \pm i\omega$) as its only eigenvalues on the imaginary axis. The fold-Hopf bifurcation is sometimes also referred to as zero-Hopf, zero-pair or Gavrilov-Guckenheimer bifurcation.

Suppose we have a n-dimensional system depending on p parameters

$$\dot{x} = f(x, \mu), \quad x \in \mathbb{R}^n, \quad \mu \in \mathbb{R}^p, \tag{3.1}$$

with fixed point $x = x_0$ at $\mu = \mu_0$ (so $f(x_0, \mu_0) = 0$). If μ_0 is a fold-Hopf bifurcation point, the eigenvalue spectrum contains n_- eigenvalues λ_i^- with negative real part, n_+ eigenvalues λ_i^+ with positive real part and finally one zero and one imaginary pair: $\mathrm{Re}\,\lambda_1^-, \ldots, \lambda_{n_-}^- < 0 < \mathrm{Re}\,\lambda_1^+, \ldots, \lambda_{n_+}^+; \lambda_1^0 = 0, \lambda_{2,3}^0 = \pm i\omega$. Clearly, $n_- + n_+ + 3 = n$. We will denote the linear eigenspaces corresponding to the three groups $\{\lambda_i^-\}, \{\lambda_i^+\}, \{\lambda_i^0\}$ by E^-, E^+ and E^0, respectively: the stable, the unstable and the *center* eigenspace. The *Center Manifold Theorem* states that there exists a (only locally defined) 3-dimensional invariant manifold W^c that is tangent to E^0 in x_0. The only interesting dynamics of the system in the neighbourhood of x_0 takes place on W^c. Away from W^c, the system is either exponentially fast repelled from W^c (along the unstable manifold) or attracted towards W^c (along the stable manifold), and thus "trivial". W^c is called the *center manifold*.

The dynamics of the system around x_0 can be described (apart from the exponentially fast repelling and/or attracting to W^c) by the dynamics on the center manifold. Thus, to understand the (local) behaviour of the system in and near a fold-Hopf bifurcation, we can restrict ourselves by looking at the center manifold dynamics. A nonlinear, parameter-dependent transformation of variables allows us to isolate the description of the system on W^c (a so-called center manifold reduction). The equations describing the evolution on W^c can be cast in a standardized form; this is called the *normal form*. We will not discuss the way to calculate the normal form and the center manifold reduction, see Kuznetsov (1995, 1999) for a detailed discussion. Instead, we only give the normal form of the fold-Hopf bifurcation:

$$\begin{aligned} \dot{y} &= \rho_1 + y^2 + s\,|z|^2 + \mathcal{O}(|y, z, \bar{z}|^4) \\ \dot{z} &= (\rho_2 + i\,\omega_1)\,z + (\theta + i\,\vartheta)\,y\,z + y^2\,z + \mathcal{O}(|y, z, \bar{z}|^4) \end{aligned} \tag{3.2}$$

Here, $y \in \mathbb{R}$ and $z \in \mathbb{C}$. ρ_1 and ρ_2 are called the unfolding parameters, ω_1, θ and ϑ are coefficients whose values depend on ρ_1 and ρ_2, and $s = +1$ or -1. We assume $\theta \neq 0$ when $\rho_1 = \rho_2 = 0$. The fold-Hopf bifurcation point is at $(\rho_1, \rho_2) = (0, 0)$. There, the normal form equations have a fixed point $y = z = 0$ with eigenvalue spectrum $0, \pm i\,\omega_1$.

Truncating the normal form after the cubic terms and transforming z to polar coordinates, $z = u\,\mathrm{e}^{\mathrm{i}\phi}$, yields

$$
\begin{aligned}
\dot{y} &= \rho_1 + y^2 + s\,u^2 \\
\dot{u} &= (\rho_2 + \theta\,y + y^2)\,u \\
\dot{\phi} &= \omega_1 + \vartheta\,y
\end{aligned}
\tag{3.3}
$$

This system is called the truncated normal form. As can be seen, the equations for y and u do not depend on ϕ. Moreover, in the neighbourhood of the fixed point, $\dot{\phi} \approx \omega_1$, due to small y. The bifurcations in the above system can therefore be understood by only studying the equations for y and u, the so-called (truncated) amplitude equations. Note that the truncated amplitude equations are $\mathbb{Z}_2$-symmetric, as they are invariant under $u \to -u$. See Kuznetsov (1995) for a discussion. This symmetry is related to the S^1-symmetry (invariance under $\phi \to \phi + \phi_c$ for arbitrary constants ϕ_c) of the truncated normal form.

Qualitatively, the bifurcation diagram near the origin in the (ρ_1, ρ_2) plane is determined by the signs of s and θ. There are four different cases, or unfolding scenarios. Which scenario applies to our situation can be deduced from the fact that in our system the Hopf-bifurcation, when it is located on the upper, stable branch, is supercritical: the stable equilibrium loses its stability while a stable periodic solution is created. It leaves the scenario with normal form coefficients $s = 1, \theta < 0$ as only possibility. This can be checked by continuation of the Hopf bifurcation using (a not yet publicly available version of) CONTENT, which calculates the normal-form coefficients when the Hopf bifurcation curve crosses a fold-Hopf point. Indeed it is found that $s = 1, \theta < 0$.

We will not discuss all four unfolding scenarios but restrict ourselves to the one relevant for this study. The bifurcation diagram for the truncated normal form in case $s = 1, \theta < 0$ is shown in figure 3.4, together with phase portraits in the (u, y) plane for the various regions of the diagram. In origin of the (ρ_1, ρ_2) plane we find the fold-Hopf point. From the origin, a saddle-node line sn ($\rho_1 = 0$) and a Hopf curve hb ($\rho_1 = -\rho_2^2/\theta^2 + o(\rho_2^2)$) emanate. For $\rho_1 > 0$ no equilibria or periodic orbits exist. When crossing sn, two equilibria are created. The invariance of the line $u = 0$ (or, equivalently, $z = 0$) in the truncated normal form guarantees the existence of a heteroclinic connection on the y-axis between the two equilibria.

When crossing hb, one of the equilibria (which one depends on the sign of ρ_2) undergoes a Hopf-bifurcation. The periodic orbit born on the Hopf curve encounters a Neimark-Sacker bifurcation (also called torus bifurcation) when crossing the line $\rho_2 = 0, \rho_1 < 0$. The Neimark-Sacker line, denoted ns, also emanates from the origin of the parameter plane. Note that in the truncated amplitude equations a Hopf-bifurcation actually appears as a fixed point moving into the $u > 0$ plane. A fixed point of the truncated amplitude equations with $u > 0$ corresponds to a periodic orbit in the truncated normal form. In the same spirit, a Neimark-Sacker bifurcation and a torus in the truncated normal form equations appear as a Hopf-bifurcation resp. a periodic orbit in the truncated amplitude equations.

The invariant torus, created when the periodic orbit crossed the ns curve, blows up and eventually touches both equilibria. This happens on the curve hc (see figure 3.4),

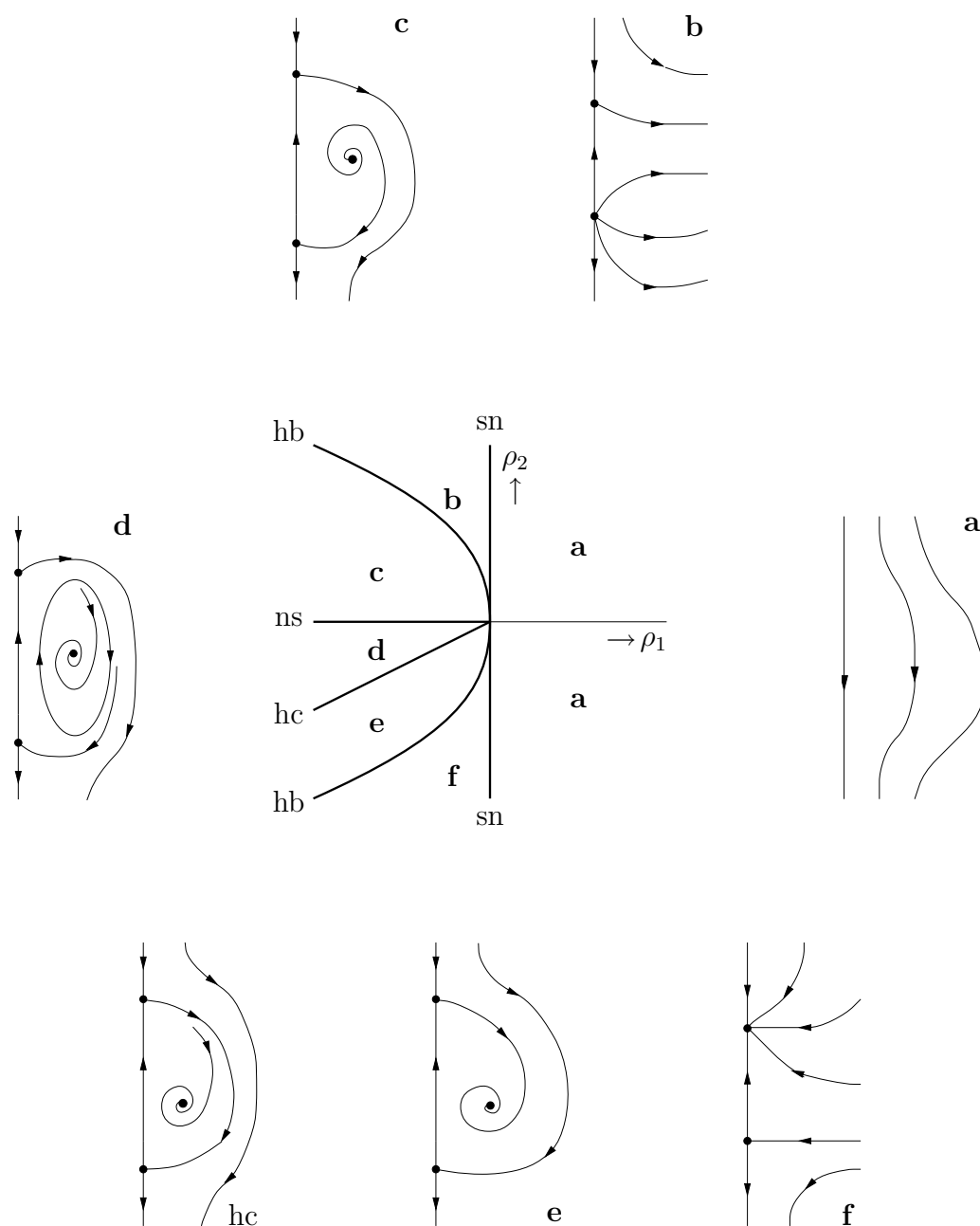Figure 3.4: *Bifurcation diagram of the truncated amplitude equations of the fold-Hopf bifurcation with $s = 1, \theta < 0$. The various lines and curves in the diagram are denoted hb for Hopf bifurcation, hc for heteroclinic cycle, ns for Neimark-Sacker bifurcation and sn for saddle-node bifurcation. The seven phase portraits* **a** *–* **f** *and* hc *show the dynamics in the (u, y)-plane.*

and creates a second heteroclinic connection between the equilibria. The second connection corresponds to a sphere-like surface in the full (truncated) normal form. Together with heteroclinic connection on the y-axis, it forms a heteroclinic cycle between the two equilibria. In figure 3.5, this cycle is drawn in three dimensions.
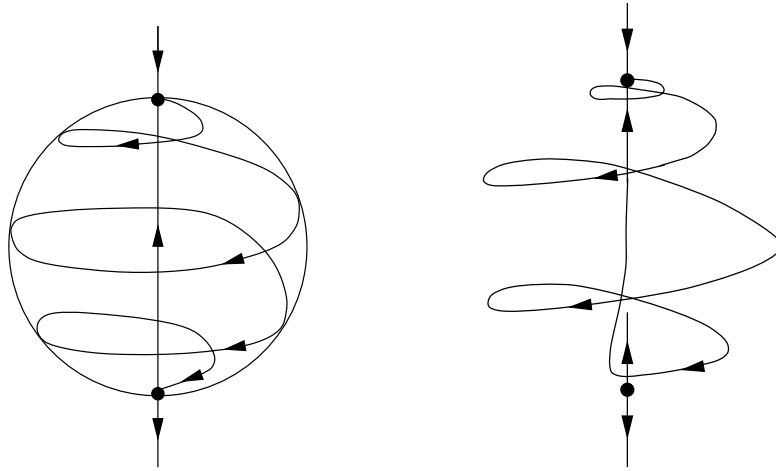

Figure 3.5: *Left: Sphere-like heteroclinic cycle between two equilibria. The cycle exists on curve hc of the truncated normal form (3.4). Right: One of the homoclinic orbits remaining after adding perturbative higher order terms to the truncated normal form.*

The bifurcation scenario sketched in figures 3.4 and 3.5 is valid for the *truncated* normal form equations (3.4). Since in practice we always deal with systems in which higher order, perturbative terms show up when carrying out the normalization, the effect of such perturbations (the $\mathcal{O}(|y, z, \bar{z}|^4)$ terms in equation (3.2)) must be considered. The perturbations do not affect the local bifurcations (saddle-node, Hopf, Neimark-Sacker) but they perturb the heteroclinic connections. For instance, the connection on the y-axis is due to the invariance of that axis, since in the truncated normal form (3.4) $\dot{u}$ is proportional to u. However, the higher order perturbations can contain terms proportional to $y^{4+l}, l \in \mathbb{N}$, that destroy the invariance and thereby the connection. In other words, the perturbations break the symmetry of the truncated normal form.

The heteroclinic cycle will, in general, be destroyed by the perturbations, leaving instead a transversal heteroclinic structure and two homoclinic orbits. These latter two are attached to either one of the two fixed points and stretch towards the other equilibrium. An example is shown in figure 3.5. The homoclinic orbit resembles the former heteroclinic cycle but does not reach all the way to the second equilibrium. The other orbit, attached to the lower equilibrium and stretching towards the upper one, is not shown. In some cases the homoclinic orbits can be of the Shil'nikov type, implying the presence of a chaotic invariant set. As the issue of chaotic behaviour is not our focus here, we do not explore this possibility.

3.4.3 Numerical bifurcation analysis

Figure 3.6 shows the results of the numerical bifurcation analysis of the model equations (3.6). The calculations were done using AUTO. The orographic height was decreased to $\gamma = 0.2$ for the calculation, corresponding to a topography amplitude of 200 meter (instead of 1 km when $\gamma = 1$). This is more realistic than $\gamma = 1$, since the individual spectral components (see section 2) of realistic topography will have amplitudes (much) smaller than the original topography itself (see e.g. Charney et al. (1981), where topographic maxima of about 2 km result in spectral amplitudes of a few hundred meters or less). Moreover, a decrease of γ will shift the region of multiple equilibria to more realistic physical values. We will come back to this in the next section. A value of 0.2 for γ results, as could have been predicted from figure 3.3, in a cusp bifurcation (c) at $(x_1^*, r) = (1.178645, -0.4965761)$, in which the two saddle-node bifurcation curves (sn1 and sn2) coincide. This implies that when $\gamma = 0.2$, only one equilibrium is found if there is no forcing in the second zonal mode (i.e. $r = 0$).

The numerical analysis recovers the theoretically predicted phenomena sketched in the previous section. The Hopf curve (hb) becomes tangent to sn1 in the fold-Hopf bifurcation point fh at $(x_1^*, r) = (0.783324, -0.821677)$. The eigenvalue spectrum at this bifurcation point is $(0, \pm i\, 0.293756, -0.103994, -0.248003 \pm i\, 0.206738)$. Thus, locally in phase space the system is attracted towards the center manifold associated with the fh bifurcation. From the fh point the Neimark-Sacker curve (ns) can be seen to originate (figure 3.6, bottom). The ns curve ends on the curve of the first period-doubling bifurcation (pd). This pd curve is not predicted by the unfolding of the fold-Hopf bifurcation, as it is not in the immediate parameter neighbourhood of the fh point. The period doubling is encountered when following the periodic orbit born on the hb curve into the parameter plane. It is the first of what is probably a cascade of period doublings. In Shil'nikov et al. (1995) and Van Veen (2002), similar (but more detailed) bifurcation analyses are presented in the context of the Lorenz 1984 baroclinic model.

As has been explained, the unfolding of the truncated normal form of the fold-Hopf bifurcation contains a heteroclinic bifurcation curve on which a heteroclinic cycle exists that connects the two equilibria eq1 and eq2. The cycle will be broken under perturbations, leaving homoclinic orbits that are nearly heteroclinic cycles. One of these orbits is found numerically by continuation of the periodic orbit created on the curve hb. Starting the continuation in r of the periodic orbit born on hb at $x_1^* = 0.9$, the orbit becomes homoclinic to the upper equilibrium eq1 at $r = -1.188582$. During the continuation, the first period-doubling curve pd is crossed twice. By continuation of the homoclinic bifurcation, a curve is obtained that winds in towards the point fh. In figure 3.6 this curve is denoted ho. Note that part of the curve falls beyond the limits of the upper figure. The two segments in the figure are connected via a turning point at $(x_1^*, r) = (1.806245, -0.5173513)$. When the curve approaches the fh point, its windings are very close to each other and calculation becomes increasingly harder. Therefore, the continuation of the ho curve doesn't reach all the way to the fh point, although that is were the curve can be expected to really end. The second homoclinic curve, not shown in figure 3.6, also winds in towards the fh point. It intersects the first homoclinic curve ho infinitely many times near fh. The orbits on this second curve are homoclinic to eq2.
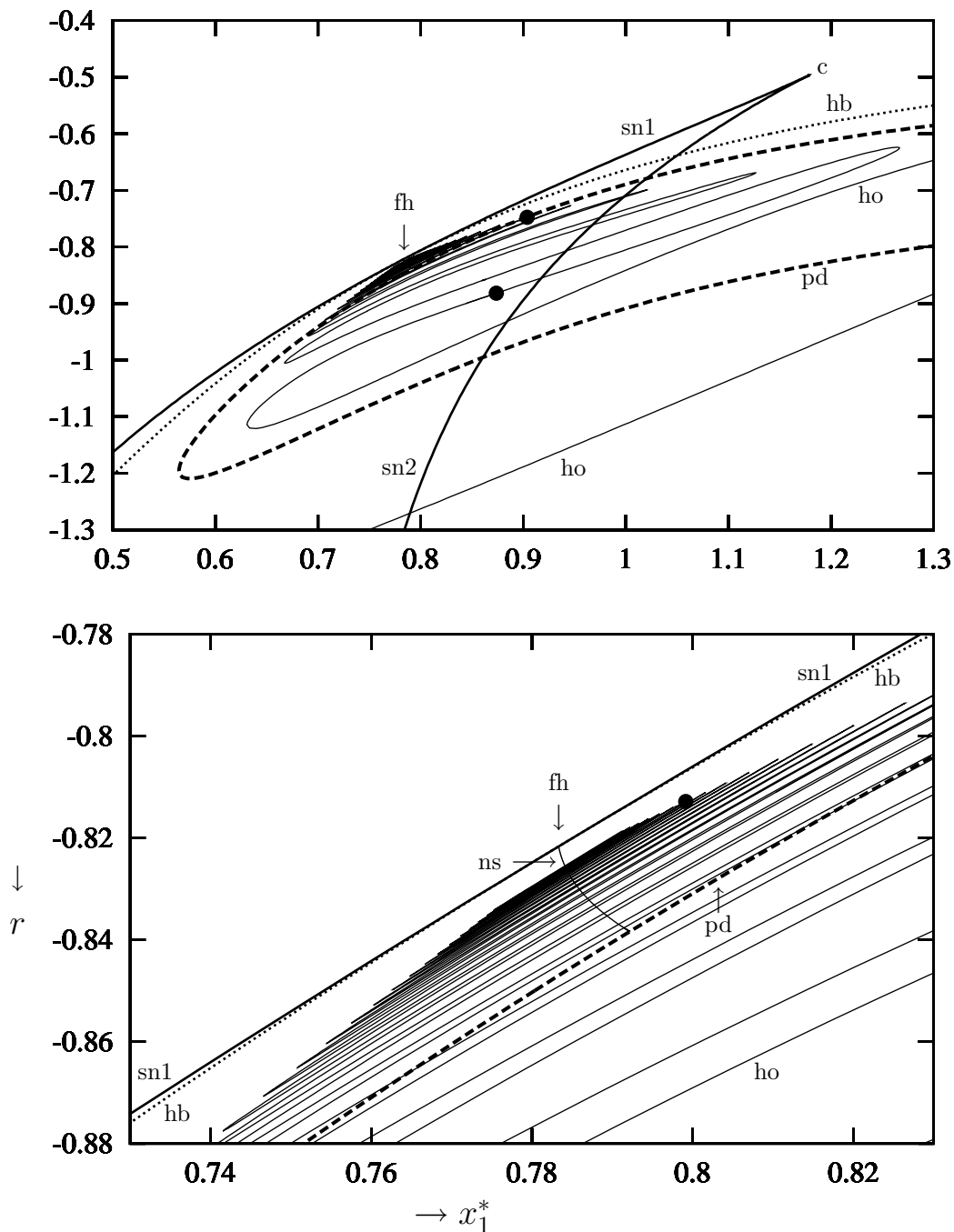
Figure 3.6: *Bifurcation diagram with fold-Hopf bifurcation (fh) as guiding center. Shown are the saddle-node bifurcation curves sn1 and sn2 (coming together in the cusp bifurcation c), the Hopf bifurcation curve hb and the period doubling curve pd. The thin solid line winding in towards the fold-Hopf point is the homoclinic bifurcation curve ho. Finally, in the magnification (bottom) the curve of the Neimark-Sacker (or torus) bifurcation ns is visible: it emanates from the fh point and ends on the pd curve. The homoclinic orbits from three locations on the curve ho, indicated by ●, are shown in figure 3.7.*

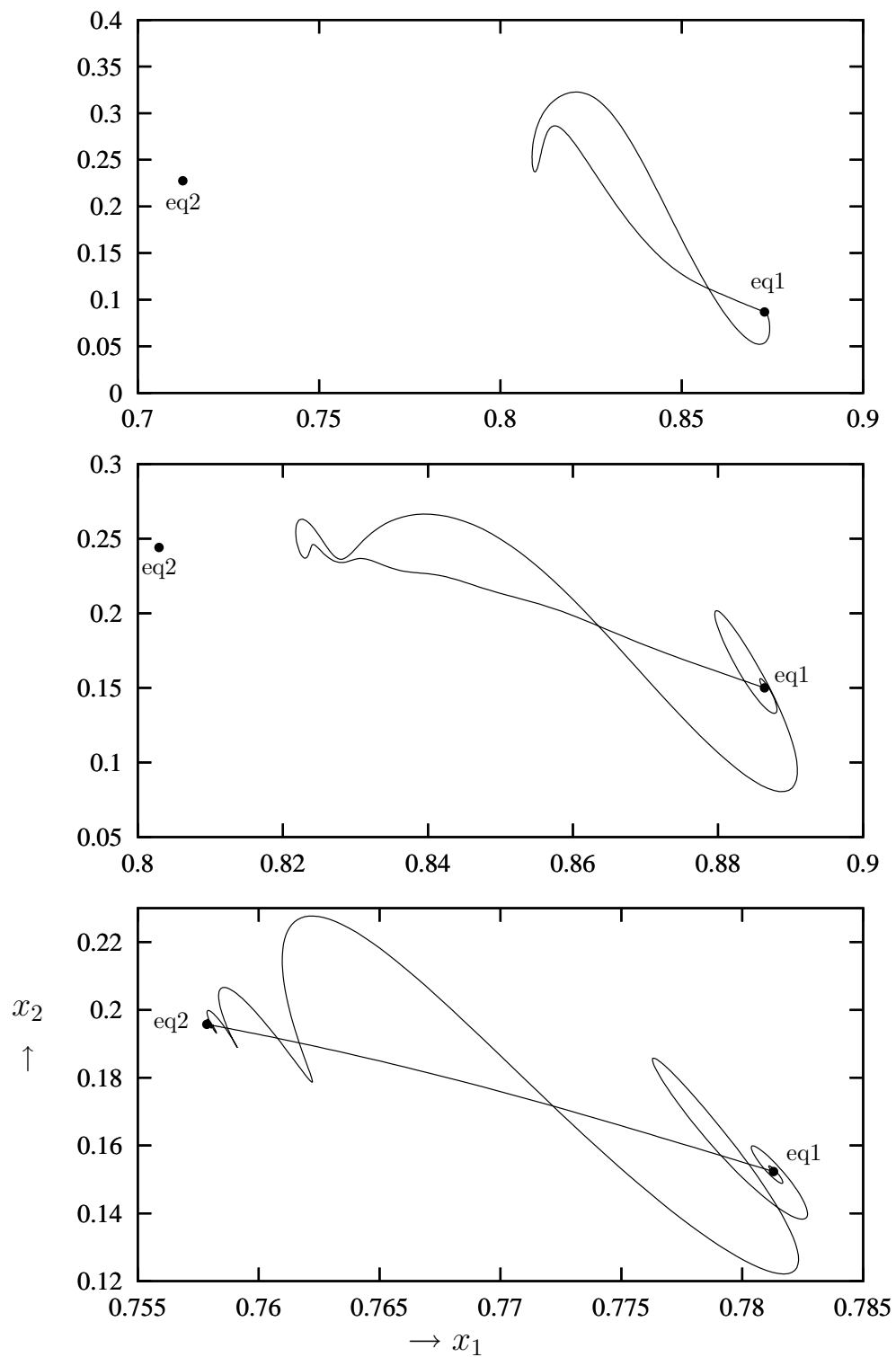Figure 3.7: *Homoclinic orbits at various points on the homoclinic curve ho. The orbits are all attached to the zonal equilibrium eq1 and can be seen to approach the wave-like equilibrium eq2 when moving along the curve ho towards the fold-Hopf point.*

In order to show that the homoclinic orbits indeed become nearly heteroclinic cycles near fh, three orbits taken from curve ho are depicted in figure 3.7 (see also figure 3.6, where the three locations on the ho curve are indicated by •). The first is the orbit that exists at $(x_1^*, r) = (0.8738674, -0.8815827)$. This is still quite far from the fold-Hopf point. The orbit is attached to the zonal equilibrium eq1 and can be seen to tend towards the second, wave-like equilibrium eq2, but doesn't come very close to it. The second orbit, with $(x_1^*, r) = 0.9037240, -0.7474002)$, tends more clearly to eq2. The third orbit, $(x_1^*, r) = 0.7991501, -0.8128589)$, comes very close to the second equilibrium and almost forms a heteroclinic cycle. In figure 3.8, this nearly heteroclinic cycle is plotted in various projections. They suggest that one part of the cycle runs over a sphere-like surface, while the other part is approximately an axis through the middle of that sphere. This is in agreement with the theoretical picture, see figure 3.5. A similar structure showed up in an EOF-based atmosphere model studied in Crommelin (2002a).

The structure of fixed points, periodic orbits and connections arising out of the fold-Hopf bifurcation induces several periodicities in the system. The unstable and stable
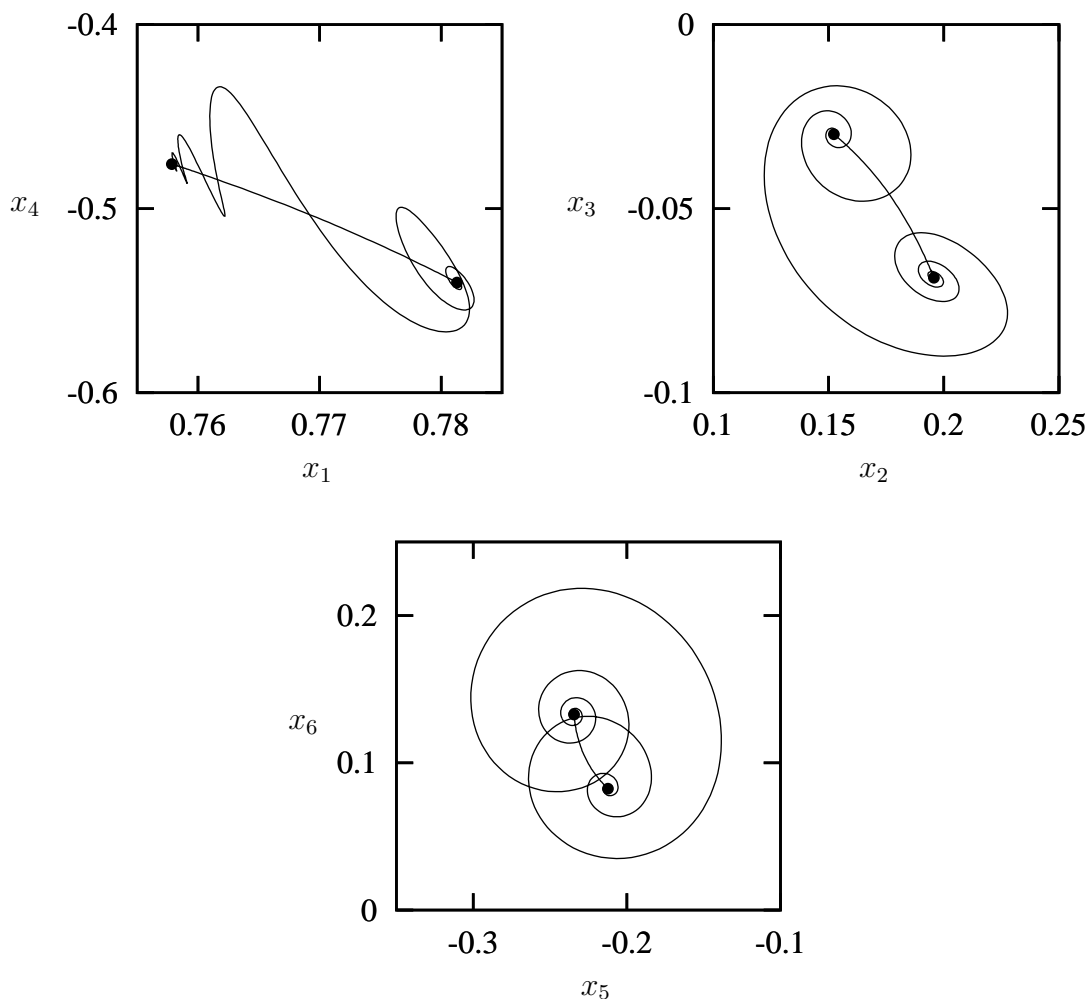

Figure 3.8: *Nearly heteroclinic cycle in variour projections.*

leading complex eigenmodes of eq1 resp. eq2 each have an oscillation period. These oscillations are not related to periodic orbits (although it is possible that in a more complex situation periodic orbits will be thrown off by eq1 or eq2, due to new Hopf bifurcations). Also, the primary periodic orbit born on the Hopf curve hb is an essential element in the unfolding of the fold-Hopf bifurcation. Furthermore, the torus created in the Neimark-Sacker bifurcation is characterized by a second period (the first being, initially, the period of the primary periodic orbit just mentioned). This second period is likely to be highly variable in parameter space, as the torus may be destroyed or touch the equilibria. Finally, the orbits homoclinic to eq1 and eq2 will give rise to either an infinity of periodic orbits (in case of a Shil'nikov type homoclinic bifurcation) or one unique periodic solution (in the other case). For mathematical details of the periodic orbits generated by homoclinic bifurcations, see Kuznetsov (1995). See also Van Veen (2002) for a detailed analysis of the relation between the homoclinic bifurcation and the period doubling cascade. In Tuwankotta (2002) heteroclinic behaviour is analysed in a general 3-dimensional system with quadratic, norm-preserving nonlinearities.

3.5   Bimodality

Numerical integration of the system, in order to see the influence of the fold-Hopf bifurcation, is complicated by the presence of the stable equilibrium eq3. In the parameter range under consideration (see figure 3.6), this fixed point doesn't become unstable. However, if we inspect the behaviour just beyond the second saddle-node curve sn2, eq3 doesn't exist anymore but its former presence still generates a stagnation point for the system. If we integrate the system at $(x_1^*, r) = (0.95, -0.801)$, a point in parameter space close to the homoclinic bifurcation curve ho and just beyond sn2 (which is at $x_1^* = 0.945$ when $r = -0.801$), regime behaviour is visible. That is, the system alternately visits the neighbourhoods of eq1 and (the former location of) eq2/eq3.

The results of an integration of 4000 time units (equivalent to 4000 days) are shown in figure 3.9. Plotting x_1 versus time, we see lingering around and transiting between eq1 on the one hand and the former location of eq2/eq3 on the other hand. Also shown is a projection of the integration orbit onto the (x_1, x_4) plane. Comparing with figure 3.8 one sees that the trajectories of the system have grown in phase space extent, but still follow roughly the same route as the nearly heteroclinic cycle in figure 3.8. The probability density function (PDF) of the distribution of states in the (x_1, x_4) plane shows two maxima towards the far ends of the elongated structure. This is the imprint of the regime behaviour on the PDF. The high phase speed of the system during its oscillating transition from the zonal to the wave-like equilibrium has the effect that states from these transitions are hardly visible in the PDF.

The system trajectory can be seen to turn around before really reaching the fixed points. It causes the PDF maxima to lie closer to the time mean state of the system than the fixed points. This has been observed elsewhere (Reinhold and Pierrehumbert, 1982; Achatz and Opsteegh, 2002): the anomaly patterns of the regimes are similar to those of the fixed points, but the amplitudes of the regime anomalies are smaller. In figure 3.10, the flow patterns corresponding to eq1 at $(x_1^*, r) = (0.95, -0.801)$ and to
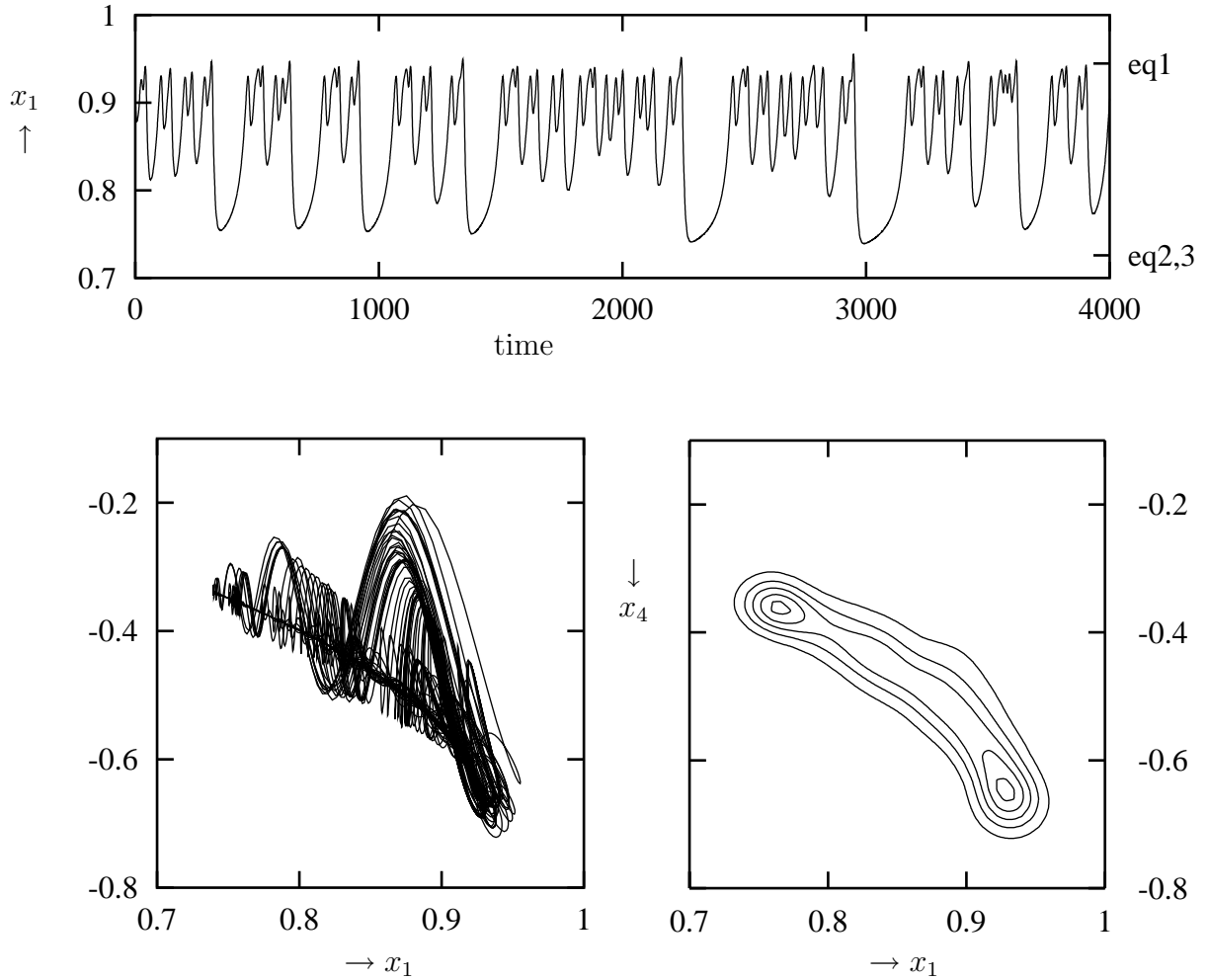
Figure 3.9: *Results of a 4000 days integration at $(x_1^*, r) = (0.95, -0.801)$. Top: x_1 versus time. The x_1-values of eq1 and of the previous location of eq2/eq3 (see text) are indicated. Bottom, left: projection of integration data onto (x_1, x_4) plane. Bottom, right: PDF in (x_1, x_4) plane calculated from integration data.*

eq2,3 at $(x_1^*, r) = (0.945, -0.801)$ (the saddle-node bifurcation point sn2, where eq2 and eq3 collide) are shown. In dimensional units (using a wind speed scaling $U_0 = 10$ m/s, related to a channel length of $2\pi \times 10^3$ km), the zonal wind speed reaches a maximum of about 30 m/s in the jet of the eq1 pattern. In the eq2,3 pattern it is about 25 m/s. These are surprisingly realistic values, especially when compared to the jet speed maxima of 150 m/s or more, usually seen for this type of model. It results from the decrease of the topographic height (0.2 instead of 1.0 km), which causes the region of parameter space with multiple equilibria to be located at more realistic values of the forcing parameters. Thus, the almost classical objection of unrealisticly strong jets, raised against CDV-like studies in e.g. Tung and Rosenthal (1985), does not hold for our choice of parameters.

To get an impression of the strength of the forcing the model is exposed to in the above integration, the forcing profile ψ^* corresponding to $(x_1^*, r) = (0.95, -0.801)$ is converted to the zonal velocity forcing profile $u = -\partial\psi^*/\partial y$. The resulting dimensional profile, using
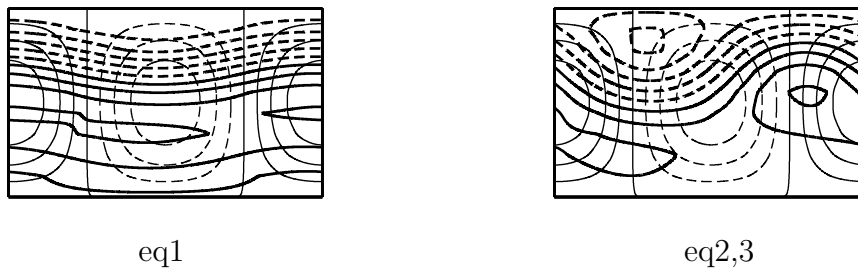
Figure 3.10: *Flow regimes corresponding to eq1 (left) at $(x_1^*, r) = (0.95, -0.801)$ and to eq2,3 (right) at the point where they collide (the saddle-node point sn2 at $(x_1^*, r) = (0.945, -0.801)$). Thick lines are streamfunction lines (contour interval 0.2 in nondimensional units), thin lines are topography contours (interval 0.05 km). Dashed contours are for negative values, solid contours for zero or positive values.*

again the scaling of $U_0 = 10$ m/s, is shown in figure 3.11. Also plotted there is the forcing profile corresponding to the fold-Hopf bifurcation point $(x_1^*, r) = (0.783324, -0.821677)$. The maximum velocity in the jet in the northern half of the domain is about 30 m/s for the forcing used in the integration, and 25 m/s for the forcing in the fold-Hopf bifurcation. These values are in the range of the velocities reached by the tropospheric jet in the real atmosphere.

In the previous section, the occurrence of several periodicities, associated to the unfolding of the fold-Hopf bifurcation, was discussed. For the integration parameter set-
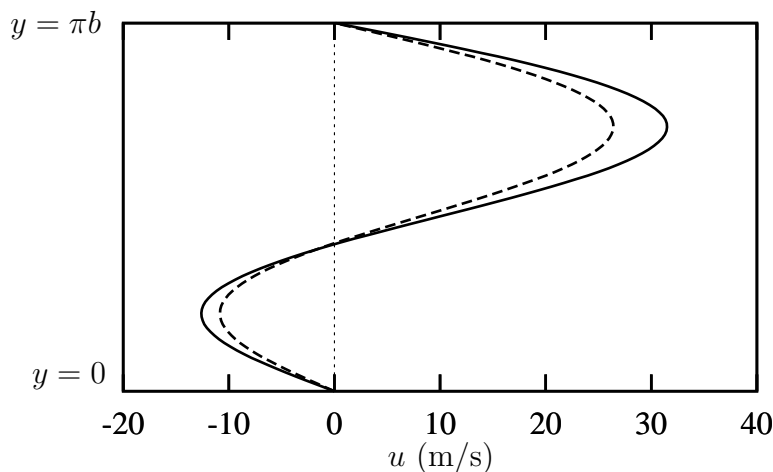

Figure 3.11: *Meridional dependence of forcing profile ψ^*, converted to zonal wind speed u in m/s. Solid line: profile corresponding to $(x_1^*, r) = (0.95, -0.801)$, used for the forward integration in section 3.5. Dashed line: profile corresponding to fold-Hopf bifurcation point at $(x_1^*, r) = (0.783324, -0.821677)$.*

tings, $(x_1^*, r) = (0.95, -0.801)$, the zonal equilibrium eq1 has a complex pair $(0.247140 \pm i\, 0.315545)$ as its only unstable eigenvalues. The oscillation period corresponding to this eigenvalue pair is about 20 days. The period related to the stable leading eigenmode of eq2 just before sn2 at $(x_1^*, r) = (0.945, -0.801)$, is about 9 days. The principal periodic orbit, born on the curve hb, has a period of 18 days at $(x_1^*, r) = (0.95, -0.801)$. The second period of the torus is not recognisable; the torus may very well not exist anymore for these parameter values. A typical period associated with the switching between the regimes does not exist anymore, due to the irregular nature of the regime switching.

3.6   Conclusion

By the simultaneous occurrence of barotropic and topographic instability in a fold-Hopf bifurcation, phase space connections are created that allow for deterministic transitions between flow regimes. The normal form equations, associated with the fold-Hopf bifurcation, describe the dynamics of the system restricted to the center manifold. Truncating the normal form equations after their cubic terms yields a system which needs the tuning of only one parameter to have a heteroclinic cycle among its solutions. The truncated normal form equations are S^1-symmetric, which explains the low codimensionality of the cycle (see Krupa (1997) for the relation between symmetries and heteroclinic cycles). This symmetry (or near-symmetry, for the non-truncated system) is made explicit by the normalisation.

The heteroclinic cycle consists of connections back and forth between two steady states. For the model studied in this paper, one of these steady states represents a situation of dominant zonal flow, the other an equilibrium flow with a mainly wave-like character. The fold-Hopf bifurcation and its unfolding thus provide a scenario in which regime transitions can be related to heteroclinic connections.

Since in a natural system perturbative terms will always be present in the model equations, the heteroclinic cycle will be broken. Two homoclinic orbits are left, each attached to one of the involved equilibria. The curves of the two homoclinic bifurcations extend quite far into parameter space. Their existence is not limited to the immediate parameter neighbourhood of the fold-Hopf point. Near the fh point the homoclinic orbits are nearly heteroclinic cycles, further away they still tend towards the unconnected second fixed point without coming really close. This dynamical configuration can explain why regime behaviour often tends to favor one regime over the other, an explanation previously speculated on in Crommelin (2002b).

The crucial ingredients for the dynamical structure described in this paper are saddle-node- and Hopf-bifurcations. The occurrence of such bifurcations are certainly no model artefacts, but on the contrary very generic features in atmosphere models. In the model studied here, the presence of topography creates saddle-node bifurcations, whereas Hopf bifurcations are the result of barotropic instability. In more complex models, the number of bifurcations will increase, not decrease, as e.g. many more topographic spectral components will be present, and the inclusion of baroclinic dynamics will add to the instabilities and thus to the number of Hopf-bifurcations. In complex models, fold-Hopf bifurcations are therefore likely to be present. The behaviour of these models will nevertheless be

complicated by many other phenomena, so the central role of fold-Hopf points in more complex models remains to be assessed. In this study, we have isolated the dynamics associated with a fold-Hopf bifurcation, in order to uncover the presence of phase space connections that can generate deterministic regime transitions.

Chapter 4

Observed non-diffusive dynamics in large-scale atmospheric flow

Abstract. The behaviour of large-scale atmospheric flow is driven both by noise, stemming mainly from small-scale, high-frequency atmospheric processes, and by the internal nonlinear dynamics of the flow system. In this study an attempt is made to isolate the latter influence from the former, in an analysis of observational data of the atmosphere system during winter. The data is projected onto the unit sphere in a reduced 3-dimensional phase space of dominant variability patterns. The probability density function (PDF) associated with the data on the unit sphere shows clear maxima that correspond with flow regimes also found in previous studies. Concepts from the theory of Markov chains and phase space partitions are used to bring out the non-diffusive, conservative dynamics of the large-scale atmospheric flow. It is found that the role of such dynamics is small but becomes increasingly important in low-pass filtered data of the atmosphere. By inspecting the asymmetries of transitions between different phase space cells, a preferred, closed path, with a preferred direction, over the unit sphere of the reduced phase space is detected. This path, or cycle, connects regimes of zonal and blocked flow in the Atlantic sector. It is reminiscent of the remnants of a heteroclinic cycle found to guide regime transitions in a model study in Crommelin (2002b).

4.1 Introduction

Among climate scientists a renewed interest is growing in the phenomenon of large scale atmospheric flow regimes. The work of Palmer (1999) and Corti et al. (1999) indicates that climate change does not so much induce new patterns of atmospheric circulation, but rather causes changes in the occupation statistics of naturally occurring circulation regimes. These regimes can capture the atmosphere system for, typically, one or two weeks. Compared with the 1950s and 60s, some regimes have been visited more frequently in the 1970s and 80s, others less frequently. This change is thought to be due to the overall change of climate over the last 50 years. Such a perspective on climate change clearly calls for a better understanding of the mechanism(s) that causes the atmosphere to switch from one circulation regime to another. It is the sensitivity of such a mechanism to changing climate parameters that will eventually determine which regime transitions will change

(become more or less likely), and thus which regimes will be visited more or less often. Unfortunately, the question what determines regime transitions is still largely unanswered.

An important contribution to the understanding of regime behaviour came from Charney and DeVore (1979, CDV from now on). They suggested that regimes correspond with equilibrium solutions of the equations describing the evolution of the large-scale flow. Their study made clear that in a severely truncated model of midlatitude flow over topography, multiple equilibria can exist that represent different flow regimes. The nonlinearity of the atmosphere system and the presence of topography make the existence of several equilibrium states and hence several regimes possible. In many different atmosphere models (often of low order) multiple equilibria were indeed found. However, a major problem was posed by the (linear) stability of these equilibria (or fixed points). Once in a stable equilibrium the system will never leave it – an unrealistic situation, as the real atmosphere continuously hops from one regime to another. The easiest way out of this problem is to associate regimes with stable (hence attracting) equilibria of the large scale flow system, and to think of transitions as being induced by noise: stochastic perturbations force the system out of the basin of attraction of one regime and into the basin of another. The noise is assumed to be caused by small scale processes in the atmosphere.

Another interpretation is possible, though. One can take the dynamical systems view of CDV one step further and interpret regimes *and* the transitions between them as deterministic instead of stochastic phenomena. Several low-order models have been shown to generate regime transitions without adding noise. Examples are the model of De Swart (1988), with 10 variables, and that of Legras and Ghil (1985) with 25 variables. Even the CDV model can be brought to regime transitions without noise when the zonal forcing profile is taken more general than in the original CDV study, see Crommelin et al. (2002). Holding on to the deterministic perspective has several advantages over taking recourse to the stochastic option. First, if regime transitions can be produced even by simple deterministic models, they can apparently be generated by internal large scale atmospheric flow dynamics, and should thus be treated as such. Second, adding noise may obscure details of regime transitions that have to do with deterministic dynamics.

In a previous study (Crommelin, 2002b) it was investigated what the deterministic nonlinear perspective can reveal when analysing the behaviour of an atmosphere model of moderate complexity without noise. With 231 variables, this barotropic model with topography and realistic forcing showed qualitatively realistic dynamics, with regime behaviour as its most prominent feature. The probability density function (PDF) of the two most dominant patterns of variability (the first two empirical orthogonal functions or EOFs) was shown to have a bimodal structure. The two PDF maxima correspond to situations of zonal resp. blocked flow. Furthermore, it was shown that the regimes can be associated with the dynamically most important steady states of the model equations. These steady states were calculated numerically.

The deterministic view leads to the interpretation of regime transitions as manifestations of *heteroclinic connections*, an interpretation also mentioned by Legras and Ghil (1985) and Kimoto and Ghil (1993b). Such connections (or orbits) are deterministic trajectories through phase space that connect one equilibrium with another. A web of these orbits, linking equilibria with each other, may very well be the dynamical structure that underlies regime behaviour. Such a structure (a so-called heteroclinic cycle) generates

regime transitions without needing noise. Moreover, the equilibria in this dynamical web should be unstable, which is the most realistic situation. In Crommelin et al. (2002) the combination of topographic and barotropic instabilities was found to be able to create such a heteroclinic cycle.

The heteroclinic connections follow certain paths through phase space, and the system can be expected to roughly follow these paths when switching from one regime to another. Since the connection from the zonal regime to the blocked regime takes a different path than the connection from the blocked back to the zonal regime, this should be visible in the behaviour of the system. The phase space region the system tends to occupy during one transition will on average be different from the region through which the transition back takes place. This was shown to be indeed the case for the moderately complex model used in Crommelin (2002b). Numerically calculated approximations of connections were found to be consistent with the average phase space preferences of the model during regime transitions.

In the present study we will investigate if similar behaviour can be detected in the real atmosphere. In order to do this we have taken the NCEP Reanalysis data (1948-2000) of the Northern Hemisphere (NH) and used daily fields for the 500 hPa geopotential height for our analysis. As is usual for studies of regime behaviour, we only look at (extended) winter data, that is, November to April.

It is to be expected that preferences and inhomogeneities in the regime transitions of the real atmosphere will not be as clear and prominent as the preferences found in e.g. the barotropic model studied in Crommelin (2002b). The low-frequency dynamics of the real atmosphere is covered under a lot of noise and thus hard to detect. We will therefore make use of an approach put forward by Pasmanter and Timmermann (2002), who use phase space partitions and Markov chains to study ENSO dynamics. This approach allows to distinguish between the purely diffusive (irreversible) dynamics in a system, which has an inherently stochastic nature, and the non-diffusive (conservative, reversible) dynamics in the system. The influence of dynamical structures such as periodic orbits, or preferred paths due to heteroclinic or homoclinic connections, should be visible in the latter part of the dynamics. We will use this approach to study the observational data of the NH flow.

In section 2 the preparation of the NCEP observational data, to make it suitable for analysis, is summarized. In section 3 a somewhat novel way to calculate the probability density function (PDF) is described. The datapoints in the normalized reduced phase space of the first three Principal Components (PCs) are projected onto the unit sphere. The PDF of the projected data is calulated on the sphere, making the non-Gaussianity of the data-distribution, and the flow regimes present, clearly visible. In section 4 some theory on Markov chains, phase space partitions and doubly stochastic matrices is reviewed. In section 5 these concepts are used to analyse the flow of the atmospheric system through the reduced phase space, or more precise, on the unit sphere. A conclusion follows in section 6.

4.2   Data

The data used are the 1948-2000 NCEP Reanalysis data for the Northern Hemisphere (17.5N-87.5N). Since we are interested in large scale flow characteristics, a resolution of $5°$ of the latitude-longitude grid will be sufficient, giving 1080 grid points in total. The 500 hPa geopotential height field (Z_{500}) is a good representative of the planetary scale midlatitude flow. As regime behaviour mainly plays a role during the winter season, we perform our analysis on the data from the winter half year, which we define to be the 181 days starting at November 1. Thus, the dataset used consists of 52 winters of each 181 days long, the first winter starting on 1 November 1948, the last winter ending on 29 April 2000. To prepare the data for analysis we remove the seasonal cycle: if $Z(d, i)$ is the field at day d of winter season i, we subtract from it the average $Z_{\text{av}}(d)$,

$$Z_{\text{av}}(d) = \frac{1}{31 \times 52} \sum_{d'=d-15}^{d+15} \sum_{i=1}^{52} Z(d', i). \tag{4.1}$$

This gives the anomaly for each day with respect to the long term mean of the 31-day time segment centered around day d of the season. At the beginning and end of each winter 15 days are lost this way, leaving us with 151 daily data points for each winter. As a last step, the data is low-pass filtered by taking n-day means (several values of n will be used). The number of filtered datapoints per winter then is $151/n$, truncated to integer value. From the resulting dataset Empirical Orthogonal Functions (EOFs) and their corresponding Principal Components (PCs) are calculated. The increasing density of gridpoints at higher latitudes is taken into account in the EOF calculation by multiplying each gridpoint value with the square root of the cosine of the gridpoint latitude. Note that no detrending was performed on the data.

In figure 4.1 the flow patterns corresponding to the first 3 EOFs calculated from 5-day means (i.e. $n = 5$) are shown, with the standard deviation of their respective PCs as amplitudes. EOFs 1 and 2 closely resemble those of Kimoto and Ghil (1993a, KG1). The amount of variance accounted for by each EOF is 10.6%, 9.1% and 7.8%, respectively. The patterns of the first three EOFs are hardly different when calculated from daily data, 3- or 10-day means, but the variance spectrum changes. The variance associated with EOFs 1, 2 and 3 together ranges from 22.0% for $n = 1$ to 31.2% for $n = 10$.

We will restrict our analysis to the reduced phase space spanned by the first three EOFs, instead of the full 1080-dimensional phase space. A justification for this restriction comes from the fact that the low-frequency behaviour of the large-scale atmospheric flow is dominated by the leading EOFs. In KG1 it is found that this low-frequency behaviour is largely captured in the leading 8 EOFs. However, as is also discussed there, the limited amount of data in practice prevents a study of the dynamics in the subspace of the first 8 EOFs. E.g., for a reliable PDF estimation in 8 dimensions, over 40000 (independent) datapoints are necessary. Even more data will be needed for a study of transitions and phase space motion in 8 dimensions. The restriction to the first 3 EOFs is a compromise between dimensional requirements on the one hand and limited sample size on the other hand. The effect of the adddition of more dimensions can only be studied thoroughly with a long model dataset.

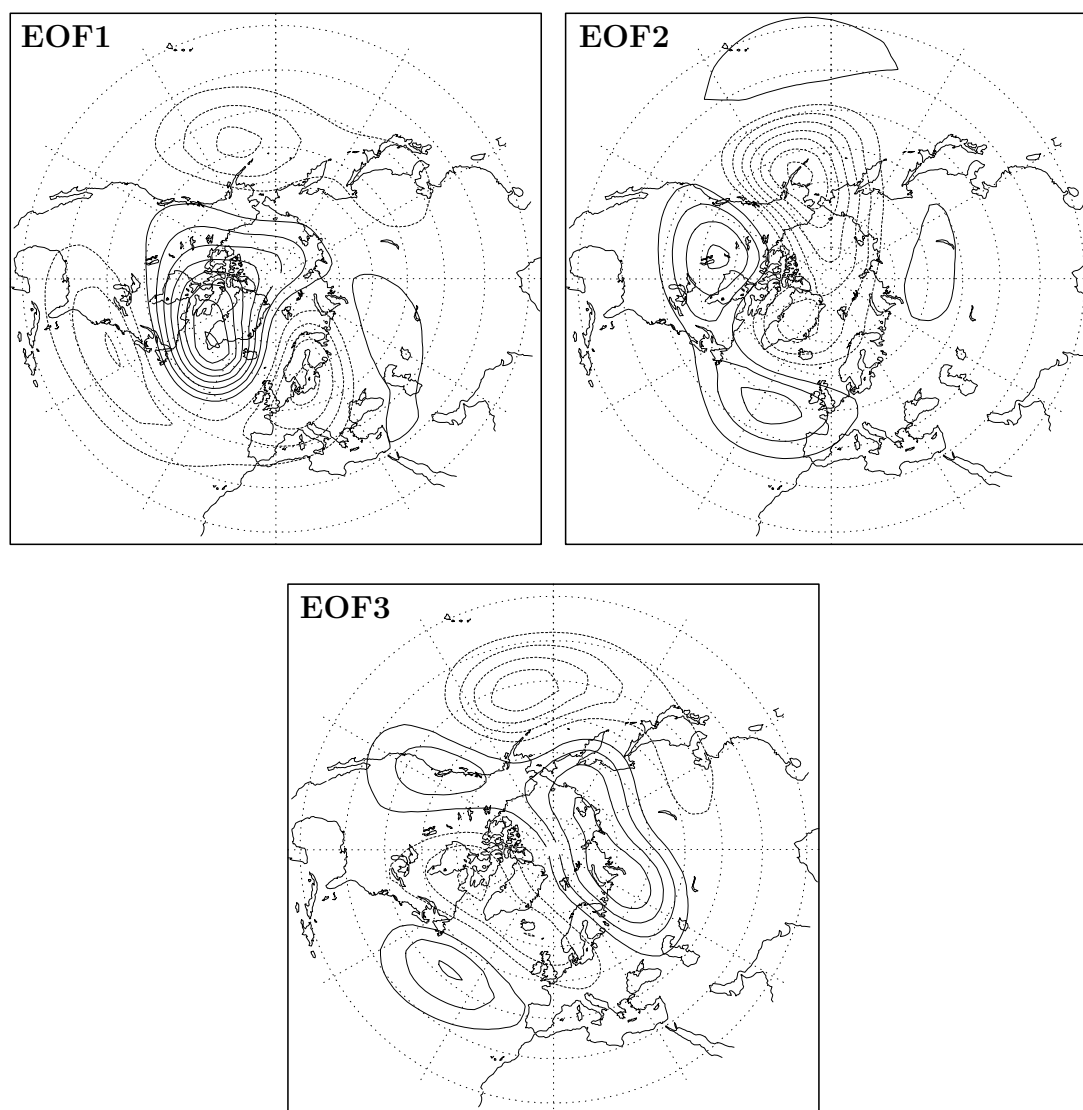Figure 4.1: *EOFS 1, 2 and 3 of the 500 mb geopotential height field, calculated from 5-day means of the 1948-2000 November to April data of the Northern Hemisphere. Associated variances: 10.6 %, 9.1 % and 7.8 %. The EOFs have been multiplied by the standard deviation of their respective PCs. Contour interval is 15 m., solid contours are positive, dashed contours negative (zero contour not drawn).*

In what follows we will mainly deal with the timeseries of the PCs normalized by their respective standard deviations. These normalized PCs will be denoted by a_1, a_2 and a_3. The corresponding normalized reduced phase space will be denoted by $\mathcal{P}$.

4.3   A spherical PDF

In studies of atmospheric regime behaviour it is rather common to focus on the projection of the data (either from observations or from model simulations) onto the plane spanned by the first two EOFs, i.e. the (a_1, a_2) plane. In this study we will use a different projection of the data onto a 2-dimensional surface: a spherical projection. This involves a coordinate transformation, from the Cartesian coordinates (a_1, a_2, a_3) to the spherical coordinates (ρ, θ, ϕ) according to

$$
\begin{aligned}
a_1 &= \rho \cos\theta \, \sin\phi \\
a_2 &= \rho \sin\theta \, \sin\phi \\
a_3 &= \rho \cos\phi
\end{aligned}
\tag{4.1}
$$

with $0 \leq \rho$, $0 \leq \theta < 2\pi$, $0 \leq \phi \leq \pi$. By projecting onto the unit sphere, more information about the structure of the flow patterns is retained than in the (a_1, a_2) projection. However, information on the amplitudes (ρ) of flow patterns is ignored: two flow patterns that have spatial correlation 1 (as far as the restriction to the first 3 EOFs is concerned), but different amplitudes, i.e. $\rho_1 \neq \rho_2$, project onto the same point (θ, ϕ). As a measure of distance between two points on the sphere we can use the angle between the two lines that connect those two points with the center point of the sphere (that is, with the origin of $\mathcal{P}$). The cosine of this angle equals the correlation between the two points.

Using the (arccosine of the) correlation as a measure of distance has been used before, e.g. by Mo and Ghil (1988) for a cluster analysis of NH atmospheric flow data, and by Kimoto and Ghil (1993b) for the calculation of 1-dimensional Probability Density Functions (PDFs) of the large-scale NH flow in the Atlantic resp. the Pacific sector. As is remarked in the latter paper, "the larger significance attached to the pattern relative to the magnitude of the anomaly is in concert with usual meteorological intuition". In this view, projection onto the unit sphere gives more information than projection onto the (a_1, a_2) plane. The calculation of a PDF on the sphere can be seen as a 2-dimensional version of the 1-dimensional angular PDFs by Kimoto and Ghil (1993b).

The spherical PDF is estimated using kernel density estimation with an Epanechnikov kernel (Silverman, 1986). The estimated PDF $f(\theta, \phi)$ is a summation of the contributions $f^i(\theta, \phi)$ from each datapoint i with coordinates $(\rho^i, \theta^i, \phi^i)$, or equivalently (a_1^i, a_2^i, a_3^i). Each contribution $f^i(\theta, \phi)$ is an Epanechnikov kernel on the sphere, centered at (θ^i, ϕ^i), with smoothing parameter σ. The total PDF $f(\theta, \phi)$ gives the probability per unit area $\sin\phi \, \mathrm{d}\theta \mathrm{d}\phi$. Thus,

$$
f(\theta, \phi) = \frac{1}{c} \sum_{i=1}^{N} f^i(\theta, \phi), \qquad f^i(\theta, \phi) = \begin{cases} 1 - (\alpha^i/\sigma)^2 & \text{if } (\alpha^i/\sigma)^2 < 1 \\ 0 & \text{otherwise} \end{cases}
\tag{4.2}
$$

in which α^i is the angle spanned by the points (θ, ϕ), the center of the sphere, and (θ^i, ϕ^i). It is easily evaluated as $\alpha^i = \arccos(\mathbf{v}^\mathrm{T}\mathbf{v}^i)$ with $\mathbf{v} = \mathbf{a}/\rho$ and $\mathbf{v}^i = \mathbf{a}^i/\rho$, the unit length
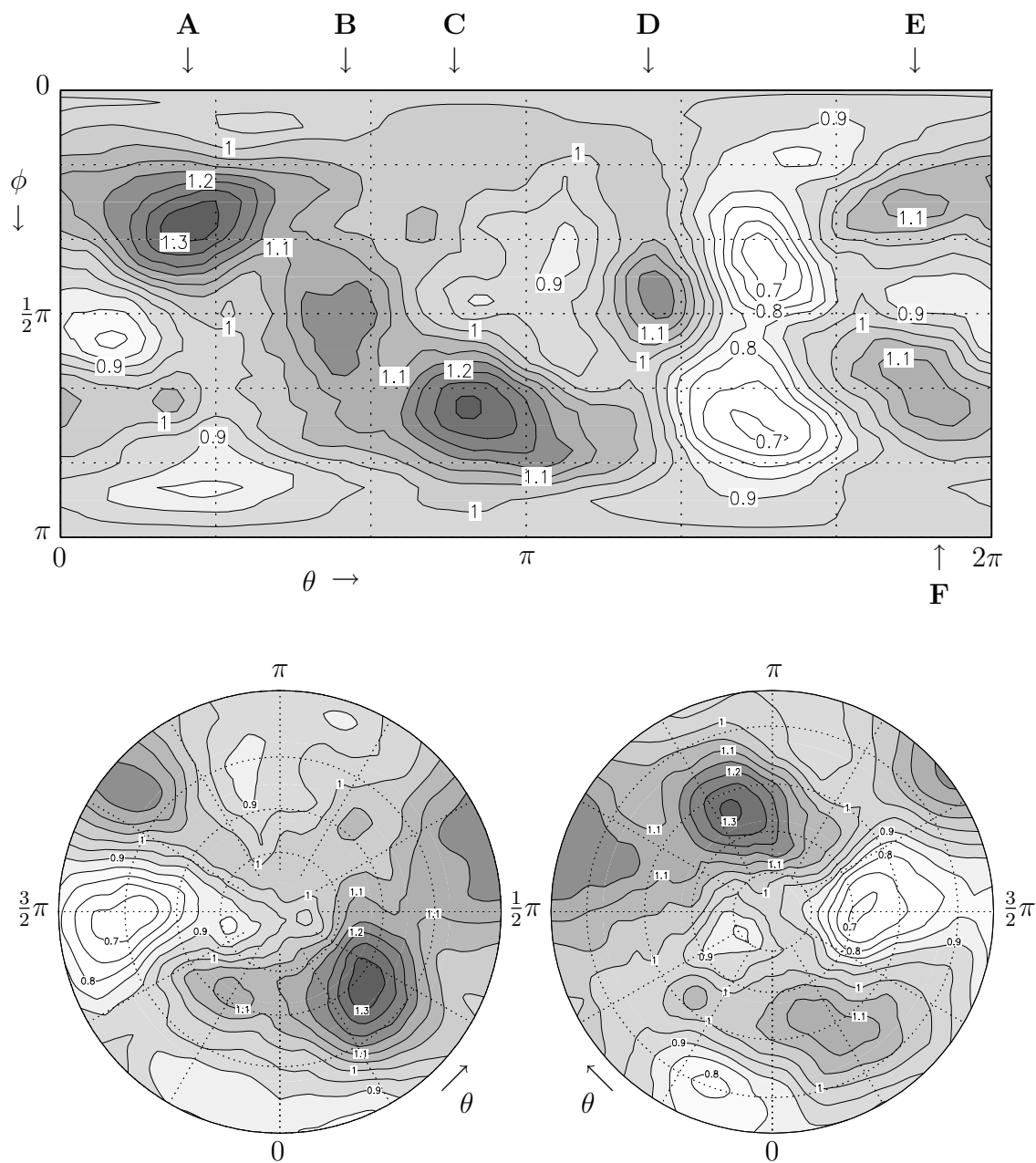
Figure 4.2: *Spherical PDF of data (5-day means) in normalized reduced phase space $\mathcal{P}$, calculated with bandwidth parameter $\sigma = \pi/6$. Shown is $f(\theta, \phi)/f_{\mathrm{G}}$, the PDF divided by the (constant) probability density $f_{\mathrm{G}}(\theta, \phi) = 1/4\pi$ of a spherically symmetric (e.g. Gaussian) distribution. A value higher (lower) than 1 indicates higher (lower) probability than Gaussian. Top: rectangular projection, $(\theta, \phi) \in [0, 2\pi) \times [0, \pi]$. Bottom: NP (left) and SP (right) stereographic projections, $(\theta, \phi) \in [0, 2\pi) \times [0, \pi/2]$ resp. $(\theta, \phi) \in [0, 2\pi) \times [\pi/2, \pi]$. Contour interval is 0.05. The letters $\mathbf{A}$-$\mathbf{F}$ indicate six main maxima in the PDF; their corresponding flow patterns are shown in figure 4.3.*

vectors in $\mathcal{P}$ pointing to (θ, ϕ) and (θ^i, ϕ^i) on the sphere. The normalisation constant c guarantees that the integral of the PDF over the sphere equals one: $\int \mathrm{d}\theta \mathrm{d}\phi \, f \sin \phi = 1$.

We do not use an adaptive kernel estimator (as is done in e.g. Kimoto and Ghil 1993a), since the projection on the sphere does not give a distribution with low-probability tails, as one would have when considering the distribution in $\mathcal{P}$. As we will see, the difference between the minimum and the maximum values of the estimated spherical PDF is less than a factor two for our choice of σ. An estimation method designed to handle long tails is therefore not considered necessary here.

In figure 4.2, the estimated PDF of the 5-day means dataset ($N = 1560$) is shown with bandwidth parameter $\sigma = \pi/6$. The value of the PDF at each point was divided by $f_{\mathrm{G}}(\theta, \phi) = 1/(4\pi)$. This is the value of the PDF all over the sphere in case of a perfectly spherically symmetric distribution (e.g. a Gaussian distribution) of datapoints in $\mathcal{P}$. Thus, it is very easy to see that the PDF in figure 4.2 is not spherically symmetric. Several areas of maximal probability can be identified, as well as regions of minimal probability. Figure 4.2 shows a "lat-lon" rectangular projection of the sphere, as well as "north pole" and "south pole" stereographic projections (NPS resp. SPS) of the upper half and the lower half of the sphere. We define the north pole of the sphere to be the point $\phi = 0$ and the south pole the point $\phi = \pi$. The NPS and SPS projections show a somewhat U-like shape of the area of increased probability, wrapped around a region of minimal probability at $\theta \approx \frac{3}{2}\pi$. This shape is reminiscent of the shape of the leading pattern resulting from the nonlinear principal component analysis performed by Monahan et al. (2001), also on NH winter data.

The number of PDF maxima depends, as usual, on the value of the bandwidth parameter. We refrain from investigating the statistical confidence levels to be associated with the various maxima, as the identification of flow regimes is not our main purpose. Whether or not the probability of an area on the sphere deviates significantly from a spherically symmetric probability distribution (e.g. a Gaussian distribution) depends both on the value of the deviation and on the magnitude of the surface of that area. We haven chosen $\sigma = \pi/6$, such that (most of) the small-scale PDF maxima have been smoothed out while still retaining some detail in the PDF. The flow patterns corresponding to the six main maxima visible in the PDF with $\sigma = \pi/6$ (figure 4.2), indicated by the letters **A**-**F**, are shown in figure 4.3. Their amplitudes correspond with points on the unit sphere in $\mathcal{P}$, i.e. $\rho = 1$. The maxima of the PDF were identified by visual inspection; no bump-hunting scheme was used.

The six flow regimes in figure 4.3 are in reasonable correspondence with the results of earlier studies, such as Mo and Ghil (1998, MG), Cheng and Wallace (1993, CW), Kimoto and Ghil (1993a, KG1), Smyth et al. (1999, SIG), Corti et al. (1999, CMP) and Monahan et al. (2001, MPF). Our regime **A** resembles the R-regime of CW, also found by MPF and SIG, the PNA (Pacific/North American) regime of KG1, and MG's cluster 1. Regime **B**, with clear zonal NAO (North Atlantic Oscillation) character, does not have a counterpart in the A-R-G classification of CW, but corresponds with the ZNAO regime of KG1. CMP's cluster A is in between our **A** and **B**. Our regime **C** can be seen to correspond well with SIG's and MPF's regime A, MG's cluster 2 and CMP's cluster B. CW's regime A looks more like our **D**. KG1's RNA pattern as well as CMP's cluster C resemble both **C** and **D**. Next, regime **E**, the reverse of **C**, has no close resemblance with
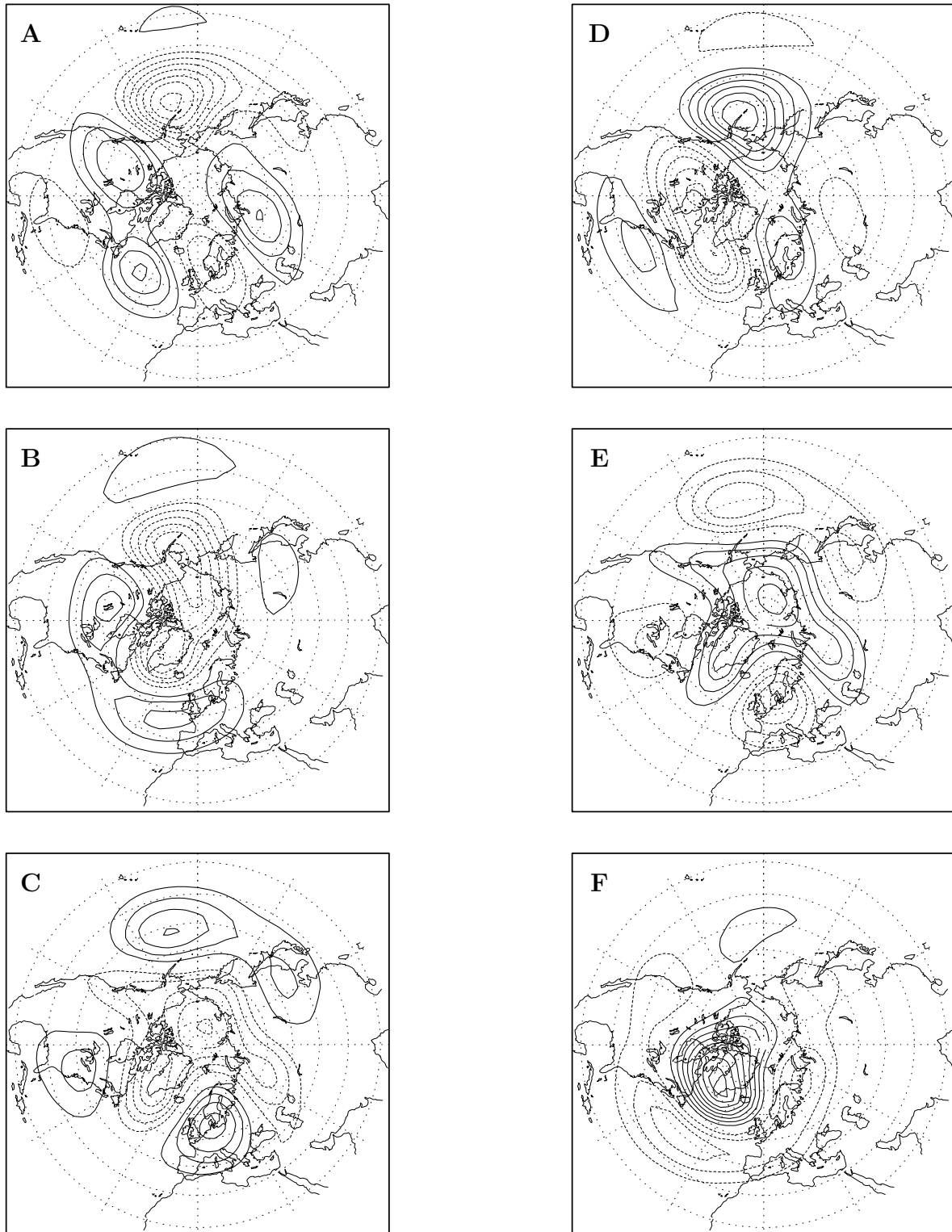
Figure 4.3: *Regime anomalies corresponding to six maxima in the spherical PDF (see figure 4.2). All maxima are taken to be on the unit sphere in $\mathcal{P}$, so $\rho = 1$. A: $(\theta, \phi) = (0.88, 0.94)$, B: $(\theta, \phi) = (1.91, 1.65)$, C: $(\theta, \phi) = (2.75, 2.22)$, D: $(\theta, \phi) = (4.00, 1.44)$, E: $(\theta, \phi) = (5.70, 0.79)$, F: $(\theta, \phi) = (5.91, 2.17)$. Contour interval is 15 m, zero contour not drawn.*

any previously found regime. Its best match would be KG1's PNA pattern. Note that it is the least prominent of our six regimes (figure 4.2). Its PDF maximum easily merges with that of **A** when increasing the bandwidth σ. Finally, regime **F** can be identified as the G- (CW, SIG, MPF) or BNAO regime (KG1), also dubbed cluster D by CMP.

Of course, for changing n and σ the results of the PDF calculation, and the corresponding regimes, will differ somewhat. For instance, **E** will merge with **A** for increasing σ and is absent when $n = 1$. Regime **D** is almost absent when $n = 10$, but prominent for lower n. Regimes **A** and **C** stay in place when changing n, whereas **F** moves towards somewhat lower θ when increasing n. Regime **B** moves to slightly higher ϕ when increasing n, and starts to merge with **C** for $n = 10$.

4.4   Phase space partitions and Markov chains

The PDF reveals the preferred flow patterns, or more generally the regions of increased and decreased probability. However, it does not tell much about the dynamical behaviour of the system under consideration, since the PDF presents a rather static view. It shows the stationary distribution of states (assuming stationarity of the system), but not the flow of the system through phase space. In order to get an idea of this phase space flow, we will divide the normalized, reduced phase space $\mathcal{P}$ in cells and consider the probabilities of transitions between the various cells.

The description of the time evolution of a system as a probability vector $p(t)$ (assigning a probability to each cell) at time t, which is multiplied by a matrix M of transition probabilities to arrive at the next timestep, i.e. $p(t + \Delta t) = M \, p(t)$, is known as a Markov chain description. It has been used in the past to model atmospheric regime behaviour by Spekat et al. (1983), Mo and Ghil (1987, 1988), Molteni et al. (1990) and Kimoto and Ghil (1993b). These studies determine a number of regimes and calculate transition matrices by counting regime transitions. During the transitions the system is assumed to be in a transient state that is not classified. Thus, not the entire (reduced) phase space is studied, but (transitions between) a few bounded regions of it. In this manner, certain preferred orders of transitions between regimes are found, see Mo and Ghil (1987, 1988) and Kimoto and Ghil (1993b).

In this section, we will not limit ourselves to certain phase space regions identified as regimes, but rather divide all of $\mathcal{P}$ into cells. The transition matrix associated with such a partition can reveal information on preferences and inhomogeneities of the flow of the system through $\mathcal{P}$. More specifically, we will partition $\mathcal{P}$ into cells all containing an equal number of datapoints, as was done in e.g. Pasmanter and Timmermann (2002). We shall refer to such a partition into cells all with an equal amount of data as an *equipartition*. The transition probability matrix M associated with an equipartition has the useful property that the sum over each row or column of the matrix equals one. Thus, if the Markov chain description is

$$p_i(t + \Delta t) = \sum_j M_{ij} \, p_j(t) \tag{4.1}$$

the matrix M satisfies

$$M_{ij} \geq 0 \ \forall\, i,j, \tag{4.2}$$

$$\sum_i M_{ij} = 1 \ \forall\, j, \tag{4.3}$$

$$\sum_j M_{ij} = 1 \ \forall\, i. \tag{4.4}$$

The first two properties are common to all transition probability matrices, not only those associated with an equipartition. Matrices satisfying them are called *stochastic* matrices. The third property is typical for matrices belonging to equipartitions. Matrices satisfying all three properties are called *doubly stochastic* matrices. A detailed treatment of equipartitions and (doubly) stochastic matrices in a geophysical context can be found in Pasmanter and Timmermann (2002, PT from now on). We will review some of the theory and results presented there.

Let us consider an equipartition of N_d datapoints into k cells with each N_c points, i.e. $N_d = kN_c$. The $k \times k$ transition probability matrix associated with this partition is again denoted by M. The k-dimensional probability vector

$$p^0 = \begin{pmatrix} 1/k \\ \vdots \\ 1/k \end{pmatrix} \tag{4.5}$$

describes the stationary probability distribution. It is called the Perron vector, and is a (right) eigenvector of M with eigenvalue 1. All other eigenvalues λ_j of M satisfy $|\lambda_j| \leq 1$. In case of M being a primitive matrix all these other eigenvalues satisfy $|\lambda_j| < 1$. If there is one unique attractor in phase space, M is irreducible and the Perron vector p^0 is unique. Thus, there is only one stationary probability distribution if M is irreducible. Moreover, in case of irreducibility p^0 is the only eigenvector with all positive elements, and all non-Perron eigenvectors either have positive and negative elements or they are complex. We will assume M to be irreducible from now on. Furthermore, in practice we will only encounter matrices M that are primitive.

The non-Perron eigenvectors, denoted by $p^j, j = 1, \ldots, k-1$, do not describe probability distributions by themselves (since their elements can be negative or complex), but rather anomalies (or deviations) from the stationary distribution p^0. Recall that the time evolution of any probability distribution is determined by repeated multiplications with M, so $p(t + n\Delta t) = M^n p(t)$. Since $M^n p^j = \lambda_j^n p^j$, there are two possibilities for the time-evolution of a probability anomaly associated with a non-Perron eigenvector p^j. If p^j and λ_j are real, the anomaly will disappear in time because $0 \leq \lambda_j < 1$. Thus, at each time step the anomaly will become smaller, but its pattern does not change. Such an eigenmode is said to be purely dissipative (or diffusive), and to have a lifetime, or e-folding time, $1/\ln|\lambda_j|$ (in time units Δt, the timestep between the consecutive datapoints). If p^j and λ_j are complex, they will always be accompanied by a complex conjugate eigenvector and eigenvalue. The probability anomaly associated with a complex pair is $\tilde{p} = c_j\, p^j + c_j^*\, p^{j*}$,

with c_j some complex constant. Writing $\lambda_j = |\lambda_j| e^{i\psi_j}$, the time evolution of the anomaly can be expressed as $M^n \tilde{p} = |\lambda_j|^n (c_j e^{in\psi_j} p^j + c_j^* e^{-in\psi_j} p^{j*})$. Thus, the time evolution of $\tilde{p}$ is a damped oscillation in the plane spanned by $\text{Re}(p^j)$ and $\text{Im}(p^j)$, with oscillation period $2\pi/\psi_j$ and, again, lifetime $1/\ln|\lambda_j|$. Obviously, the oscillation period can never be smaller than the unit time step (reflected in the fact that $\psi_j \in [0, 2\pi]$). For a reasonable description of the oscillation the period should of course be several times larger than the unit time step.

Let us formulate a null-hypothesis for the transition matrix M and thereby for the system flow through $\mathcal{P}$. Our null-hypothesis will be the assumption that M is symmetric. It corresponds to the situation in which the system described by the Markov chain is in detailed balance. This means that each transition $i \to j$ is as probable as its reverse transition $j \to i$, i.e. $M_{ji} p_i^0 = M_{ij} p_j^0$ (no summation), where p^0 is the stationary probability distribution. If the partition of $\mathcal{P}$ is an equipartition, detailed balance takes the form $M_{ij} = M_{ji}$, since in an equipartition $p_i^0 = p_j^0 \; \forall i, j$ by construction.

Clearly, it is the antisymmetric part of M that determines to what extent the null-hypothesis will be violated. If M is symmetric there will be no net probability current. In PT, the matrix M is split up in its symmetric part M^S and antisymmetric part M^A in order to distinguish between the diffusive or dissipative part of the dynamics (called irreversible) and the non-diffusive or conservative part (called reversible). M^S is again doubly stochastic, M^A is not. A system with purely dissipative dynamics has zero M^A. In contrast, M^S is still nonzero in case of purely conservative dynamics. PT calculate $\langle \Delta I \rangle$, the average information loss, or entropy production, associated with the system $p(t + \Delta t) = M p(t)$. Two contributions to the information loss can be distinguished:

$$
\begin{aligned}
\langle \Delta I \rangle &= -\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} M_{ij} \ln M_{ij} \;\; = \langle \Delta I \rangle_S + \langle \Delta I \rangle_{A/S}, \\
\langle \Delta I \rangle_S &= -\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} M_{ij}^S \ln M_{ij}^S, \\
\langle \Delta I \rangle_{A/S} &= -\frac{1}{k} \sum_{i=1}^{k} \sum_{j>i}^{k} M_{ij}^S \ln \left(1 - \left(\frac{M_{ij}^A}{M_{ij}^S} \right)^2 \right) + M_{ij}^A \ln \left(\frac{M_{ij}^S + M_{ij}^A}{M_{ij}^S - M_{ij}^A} \right).
\end{aligned}
\tag{4.6}
$$

$\langle \Delta I \rangle_S$ is the information loss one would have in a purely diffusive system $p(t + \Delta t) = M^S p(t)$. Both $\langle \Delta I \rangle \geq 0$ and $\langle \Delta I \rangle_S \geq 0$. In contrast, $\langle \Delta I \rangle_{A/S} \leq 0$, so the total information loss in a system is decreased by the presence of conservative dynamics. The most extreme example of a conservative system (the case in which M is a permutation matrix) is characterised by $\langle \Delta I \rangle = 0$, even though M^S is nonzero. $\langle \Delta I \rangle$ is bounded from above: $\langle \Delta I \rangle \leq \ln(k)$. The upper bound is reached when the equipartition is such that $M_{ij} = 1/k, \; \forall i, j$. The information $I(t)$ of a probability distribution $p(t)$ lies in its deviation from the stationary distribution p^0,

$$
I(t) = \sum_{i=1}^{k} p_i(t) \ln(p_i(t)/p_i^0),
\tag{4.7}
$$

and gets lost because repeated multiplications by M make $p(t)$ evolve in time towards p^0. The information is bounded: $I(t) \leq \ln(k)$. If the system is with certainty in cell j, so that $p_i(t) = \delta_{ij}$, the information is maximal: $I(t) = \ln(k)$.

The statistical significance test for regime transition matrices described by Vautard et al. (1990) is not useful for the transition matrices we are studying here. As was pointed out in that paper, the fact that some regimes have higher probability than others makes the assumption that all transitions are equally probable unfit as a null-hypothesis to test the transition matrix. As an alternative Vautard et al. present a way to test statistical significance, using Monte-Carlo simulations, that takes into account the unequal regime sizes. When using, as we do, an equipartition, the problem of unequal regime (or cell) sizes disappears. Moreover, even with cells of equal size the assumption of equally probable transitions is far from what we can expect. Since we partition all of $\mathcal{P}$, including the regions of transient motion, the system will usually not jump from one cell to another all across phase space (unless it is close to the origin of $\mathcal{P}$). The vast majority of transitions will be to neighbouring cells (or from a cell back to itself), maybe some will skip one or two cells. Testing the hypothesis of equally probable transitions between all cells is therefore hardly informative in our situation, as it will certainly be rejected.

Nevertheless, another significance test is possible. If the assumption that the flow through $\mathcal{P}$ is purely diffusive (irreversible) is correct, pairs of transitions $i \to j$ and $j \to i$ should be equally probable, so the transition probability matrix M should be symmetric. Therefore, it makes sense to test whether the nonzero elements of M^A are statistically significant.

From the equipartition of $\mathcal{P}$ in k cells with each N_k datapoints the transition matrix T can be calculated, counting the number of transitions between cells. The transition probability matrix M is derived from T by $M = N_k^{-1} T$. Their respective symmetric and antisymmetric parts have the same relation: $M^S = N_k^{-1} T^S$, $M^A = N_k^{-1} T^A$. If the system dynamics is driven by M^S, we expect the N_k transitions from cell j to other cells to result in $T_{ij}^S = N_k M_{ij}^S$ transitions to cell i. T_{ij}^S is the expectation value from N_k independent binomial Bernoulli trials with at each trial a probability M_{ij}^S of "success" (transition to i) and a probability $1 - M_{ij}^S$ of "no success" (transition to any other cell). The standard deviation associated with the binomial distribution is $S_{ij} = (M_{ij}^S(1 - M_{ij}^S)N_k)^{1/2}$. For large enough N_k the binomial distribution can be approximated by the normal distribution (Feller, 1968). T_{ij} will therefore with 68.3% certainty lie in the interval $(T_{ij}^S - S_{ij}, T_{ij}^S + S_{ij})$, with 95.4% certainty in the interval $(T_{ij}^S - 2 S_{ij}, T_{ij}^S + 2 S_{ij})$, etcetera.

Since by construction $T_{ij}^A = -T_{ji}^A$, we only have to calculate the significance of T_{ij}^A in order to assess the significance of the asymmetry inherent to a transition pair $i \to j$, $j \to i$. T_{ij} lies with 95% certainty between $T_{ij}^S \pm 1.96\, S_{ij}$, so T_{ij}^A is significant at the 95% confidence level if $|T_{ij}^A| > 1.96\, S_{ij}$. To be significant at the 90%, 80% and 70% confidence level, $|T_{ij}^A|$ must exceed $1.64\, S_{ij}$, $1.28\, S_{ij}$ resp. $1.04\, S_{ij}$. Note that it is T^A rather than M^A which is tested, since a small but nonzero element of M^A may very well be significant if N_k is large enough, whereas a large element of M^A can be insignificant if N_k is small.

4.5 Results

4.5.1 Conservative versus diffusive dynamics

In order to make use of the special properties of doubly stochastic matrices, described in the previous section, we will divide $\mathcal{P}$ in such a way that we have an equipartition. Regarding k, the number of cells, PT use as a rule of thumb $k^2 \leq N_d$, so that in principle each transition can occur at least once. From the 5-day means we have $N_d = 1560$, so $k \leq 39$. In general, the lower k, the lower the information loss $\langle \Delta I \rangle$ but also the larger the ratio between $\langle \Delta I \rangle$ and $\langle \Delta I \rangle_S$, indicating a more dominant role of the diffusive part of the dynamics. PT present $\mu = 1 - k^{-1} \sum_{i,j} |M_{ij}^A|$ as a measure of the purely diffusive dynamics inherent to M, ranging from 1 (totally diffusive, all M_{ij}^A are zero) to 0 (totally conservative, M is a permutation matrix). It can be shown that $1 - \mu$ is a measure for the fraction of non-symmetric transitions: $1 - \mu = (\sum_{i,j} |T_{ij}^A|)/(\sum_{i,j} T_{ij})$. The measure μ is also seen to increase for lower k.

We will partition $\mathcal{P}$ such that the cell boundaries only depend on θ and ϕ. The partition can be drawn on the unit sphere on which the PDF was calculated, thereby enabling an easy visualisation and interpretation of probability flows. First we draw boundaries of constant ϕ, dividing the unit sphere in $\mathcal{P}$ into zonal bands. These bands are then each divided into cells, by drawing boundaries of constant θ within each band (starting with $\theta = 0$). More cells are created when the band is near $\phi = \pi/2$; the zonal

n	k	geometry of partition	N_k	$\langle \Delta I \rangle$	$\langle \Delta I \rangle_S$	$\langle \Delta I \rangle_{A/S}$	μ
1	12	1-2-6-2-1	654	1.25	1.25	-0.006	0.95
1	24	1-6-10-6-1	327	1.54	1.55	-0.01	0.93
1	36	1-7-10-10-7-1	218	1.76	1.79	-0.03	0.89
1	48	1-9-14-14-9-1	163	1.91	1.95	-0.04	0.87
1	84	1-15-26-26-15-1	93	2.28	2.35	-0.07	0.79
3	12	1-2-6-2-1	216	1.90	1.92	-0.02	0.89
3	24	1-6-10-6-1	108	2.31	2.36	-0.05	0.81
3	36	1-7-10-10-7-1	72	2.53	2.63	-0.10	0.74
3	48	1-9-14-14-9-1	54	2.68	2.83	-0.15	0.67
5	12	1-2-6-2-1	130	2.12	2.14	-0.02	0.86
5	24	1-6-10-6-1	65	2.55	2.64	-0.09	0.74
5	36	1-7-10-10-7-1	43	2.71	2.89	-0.18	0.61
10	12	1-2-6-2-1	65	2.28	2.33	-0.05	0.78
10	24	1-6-10-6-1	32	2.63	2.79	-0.16	0.64

Table 4.1: *Examples of different equipartitions of $\mathcal{P}$. Data from n-day means are divided in k cells with each N_k datapoints. The total information loss $\langle \Delta I \rangle$ is the sum of the information loss $\langle \Delta I \rangle_S$, due to diffusive dynamics, and the (negative) information loss $\langle \Delta I \rangle_{A/S}$, produced by conservative dynamics. μ is a measure of the relative importance of the diffusive dynamics, see text.*

bands at the top and the bottom of the sphere will not be divided anymore, but they will each be one cell. E.g., we draw boundaries of constant ϕ to divide zonal bands with N_k, $6N_k$, $10N_k$, $6N_k$ and N_k datapoints each, and subsequently divide these bands into 1, 6, 10, 6 and 1 cell, respectively. The total number of cells is k, the number of datapoints used is kN_k. A possible remainder of $N_d - kN_k$ points at the end of the dataset is not used. Once the partition is made, M is calculated by counting transitions. Transitions from the last datapoint of one winter to the first datapoint of the next winter are left out, thereby introducing an unavoidable but (very) small deviation of M from the precise properties of doubly stochastic matrices (equations (4.3) and (4.4)). This deviation will be ignored.

In table 4.1 results are given for several examples of partitions of the daily data ($N_d = 7852$), the 3-day means ($N_d = 2600$), the 5-day means ($N_d = 1560$) and the 10-day means ($N_d = 780$). They show that the conservative dynamics become increasingly important for low-pass filtered data. This is in agreement with the general intuition that high-frequency motion in the atmosphere is very noisy, but low-frequency dynamics are more structured and less noisy. For instance, a 1-6-10-6-1 partition of the 10-day means data, i.e. a partition in $k = 24$ cells, gives $\mu = 0.64$ whereas the same partition gives $\mu = 0.93$ for the daily data. Even a 1-15-26-26-15-1 partition of the daily data still gives $\mu = 0.79$. In contrast, the finest partitions of the 5- and 10-day means can make μ drop to 0.61 resp. 0.64 (it must be admitted that the statistical significance of these differences has not been tested).

4.5.2 Significant transition asymmetries

By inspecting M^A (or equivalently T^A), it is possible to get an impression of the net probability current at each timestep. Significant inhomogeneities and structures in the flow of the system through $\mathcal{P}$ should be visible in this current, provided the partition of $\mathcal{P}$ allows for these structures to be seen. Also, if the timestep of the dataset is too large, the sampling may be too coarse to follow the flow induced by the conservative dynamics underlying M^A. This makes the use of the 10-day means less suitable.

In the previous section, a way to assess the statistical significance of the elements of M^A or T^A was discussed. We still have to decide which number of datapoints can be considered large enough for the normal distribution to be an acceptable approximation of the binomial distribution. We will choose $T_{ij}^S \geq 10$, i.e. the number of symmetric transitions $j \rightarrow i$ must be at least 10 for T_{ij}^A to be potentially significant at all. This will in practice limit the number of cells, k. We will use the daily data in order to be able to find significant asymmetries without having to decrease k too much. In that way the resolution of the partition will not be too coarse.

We have partitioned the daily data in three different ways (1-4-6-4-1, so $k = 16$, $N_k = 490$; 1-4-6-6-4-1, with $k = 22$, $N_k = 356$; 1-4-10-10-4-1, with $k = 30$, $N_k = 261$) and analysed the transition asymmetries by inspecting the associated matrices T^A. The results are shown in figure 4.4. The cells, as well as the significant transition asymmetries, are drawn in the (θ, ϕ)-plane of the rectangular projection of the sphere. Transition asymmetries (i.e. elements from T^A) significant at a confidence level of at least 70% are drawn solid, transition asymmetries just below 70% but occurring often ($T_{ij}^S \geq 20$) are
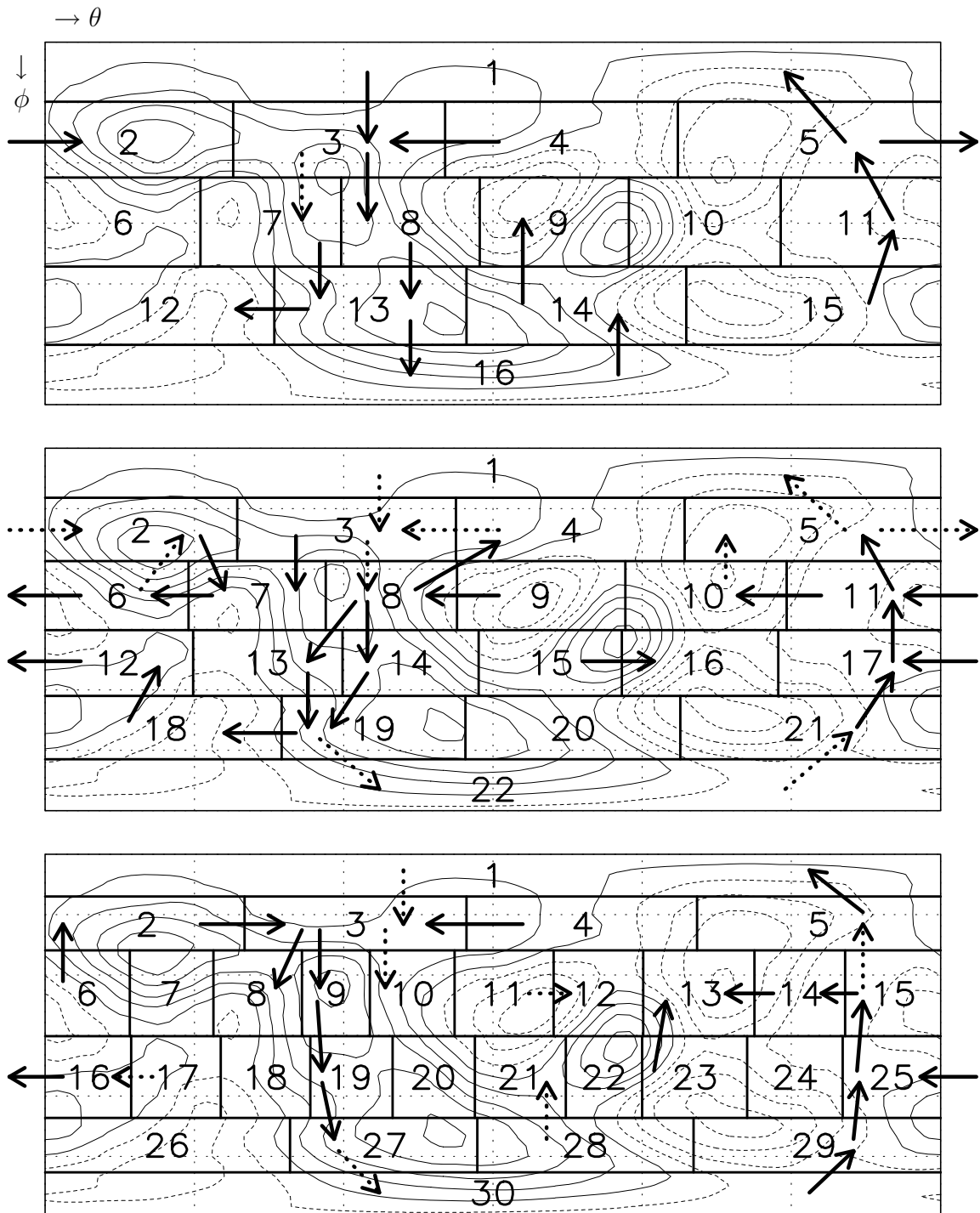
Figure 4.4: *Three different equipartitions of the daily data, in 16 (top), 22 (middle) and 30 (bottom) cells. Solid arrows indicate transition asymmetries significant at 70% or more. Dotted arrows indicate asymmetries occurring frequently ($T_{ij}^S \geq 20$) but just below 70% significance. The partitions show a preferred path of transition asymmetries, from $\phi \approx 0$ along $\theta \approx 0.67\,\pi$ to $\phi \approx \pi$ and back along $\theta \approx 1.9\,\pi$ to $\phi \approx 0$. Thin lines are contours of the scaled spherical PDF $f(\theta, \phi)/f_{\mathrm{G}}$ of the daily data (contour interval 0.05, solid contours for values ≥ 1, dashed contours for values < 1).*

dotted. To make comparisons with the PDF for the daily data easy, the contours of the scaled spherical PDF $F(\theta, \phi)/f_G$ of the daily data are also drawn (cf. figure 4.2).

A preferred path of transition asymmetries can be seen in the partitions of figure 4.4: from $\phi \approx 0$ along $\theta \approx 0.67\,\pi$ to $\phi \approx \pi$ and back along $\theta \approx 1.9\,\pi$ to $\phi \approx 0$. Some segments of this path can be just below 70% significance, and one segment is missing in the $k = 16$ partition. However, all segments are significant in at least one of the three partitions shown. Moreover, some segments are highly significant: $15 \rightarrow 11 \rightarrow 5$ at 90% for $k = 16$ and $17 \rightarrow 11$ at 95% for $k = 22$. For $k = 30$ transitions $19 \rightarrow 27$ and $25 \rightarrow 15$ are asymmetric at 95% significance, and $9 \rightarrow 19$ and $30 \rightarrow 29$ at 90%. The fact that the same path is found in all three partitions (as well as in other partitions not shown) also illustrates its robustness.

4.5.3 A preferred cycle connecting regimes

In figures 4.5 and 4.6 the evolution of the atmospheric flow along the preferred path is presented. The flow patterns are 500 mb geopotential height anomalies, with $\rho = 1$ chosen as amplitude. The evolution shows an alternation between episodes of mainly zonal anomalies ($\phi \approx 0.5\pi$) and episodes of more meridionally oriented anomalies ($\phi \approx 0, \pi$). The negative anomalies over the North Atlantic and the Pacific, present when $\phi = 0$, start to merge, and form one large negative arctic anomaly when $\phi \approx 0.5\pi$ (and $\theta = 0.67\pi$). Further down the path, this negative anomaly gets elongated meridionally over the Eurasian continent and the westcoast of North America. Over the North Atlantic and the North Pacific, it is replaced by positive anomalies. Reaching the bottom of the unit sphere, $\phi = \pi$, the pattern consists again of mainly meridionally oriented anomalies. The second half of the cycle, along $\theta = 1.9\pi$, shows a reverse evolution: positive North Atlantic and Pacific anomalies merge, forming a positive arctic anomaly with negative anomalies surrounding it ($\phi \approx 0.5\pi$). In the last part of the cycle, the arctic anomaly again stretches southward, and at $\phi = 0$ the anomaly pattern is back in its meridional state.

The preferred path can be seen to pass along several flow regimes. At $(\theta, \phi) = (0.67\,\pi, 0.35\,\pi)$ it leads through the zonal NAO regime, and at $(\theta, \phi) = (0.67\,\pi, 0.75\,\pi)$ it comes near the A regime of the CW classification. Somewhat further, at $(\theta, \phi) = (1.9\,\pi, 0.6\,\pi)$ the path encounters the G or blocked NAO regime. The regimes of zonal and blocked NAO have a mainly zonally symmetric character, in particular the latter.

One can wonder whether the temporal evolution along the path slows down in or near the regimes. This would be reflected in the diagonal matrix elements M_{ii}, as they give the probabilities of transitions from each cell back to itself. If the atmosphere tends to persist for a while in a certain regime, the corresponding cell (or cells) is likely to have a relative large probability of transitions back to itself. A problem here is posed by the fact that the regimes can be spread out over several cells. Nevertheless, for the three partitions shown in figure 4.4, the corresponding transition probability matrices have maximal diagonal elements for cells in or near the blocked NAO regime. For the $k = 16$ partition, $0.55 \leq M_{ii} \leq 0.64$, with $M_{10,10} = M_{11,11} = 0.64$ and $M_{5,5} = M_{15,15} = M_{16,16} = 0.63$. The $k = 22$ partition gives $0.50 \leq M_{ii} \leq 0.60$, with $M_{11,11} = M_{16,16} = 0.60$, $M_{21,21} = 0.59$ and $M_{10,10} = M_{17,17} = 0.58$. The $k = 30$ partition has a clear maximum, $M_{25,25} = 0.57$, with

$\phi = 0$ $\qquad\qquad$ $\phi = 0.1\,\pi$ $\qquad\qquad$ $\phi = 0.2\,\pi$

$\phi = 0.3\,\pi$ $\qquad\qquad$ $\phi = 0.4\,\pi$ $\qquad\qquad$ $\phi = 0.5\,\pi$

$\phi = 0.6\,\pi$ $\qquad\qquad$ $\phi = 0.7\,\pi$ $\qquad\qquad$ $\phi = 0.8\,\pi$

$\phi = 0.9\,\pi$ $\qquad\qquad$ $\phi = \pi$

Figure 4.5: *Evolution of the atmospheric flow (500 mb geopotential height anomaly) during the first half of the preferred path through the reduced phase space $\mathcal{P}$ (see figure 4.4). Spherical coordinates are $\rho = 1$, $\theta = 0.67\,\pi$ and ϕ varying from 0 to π in steps of $0.1\,\pi$. Contour interval is 15 m.*
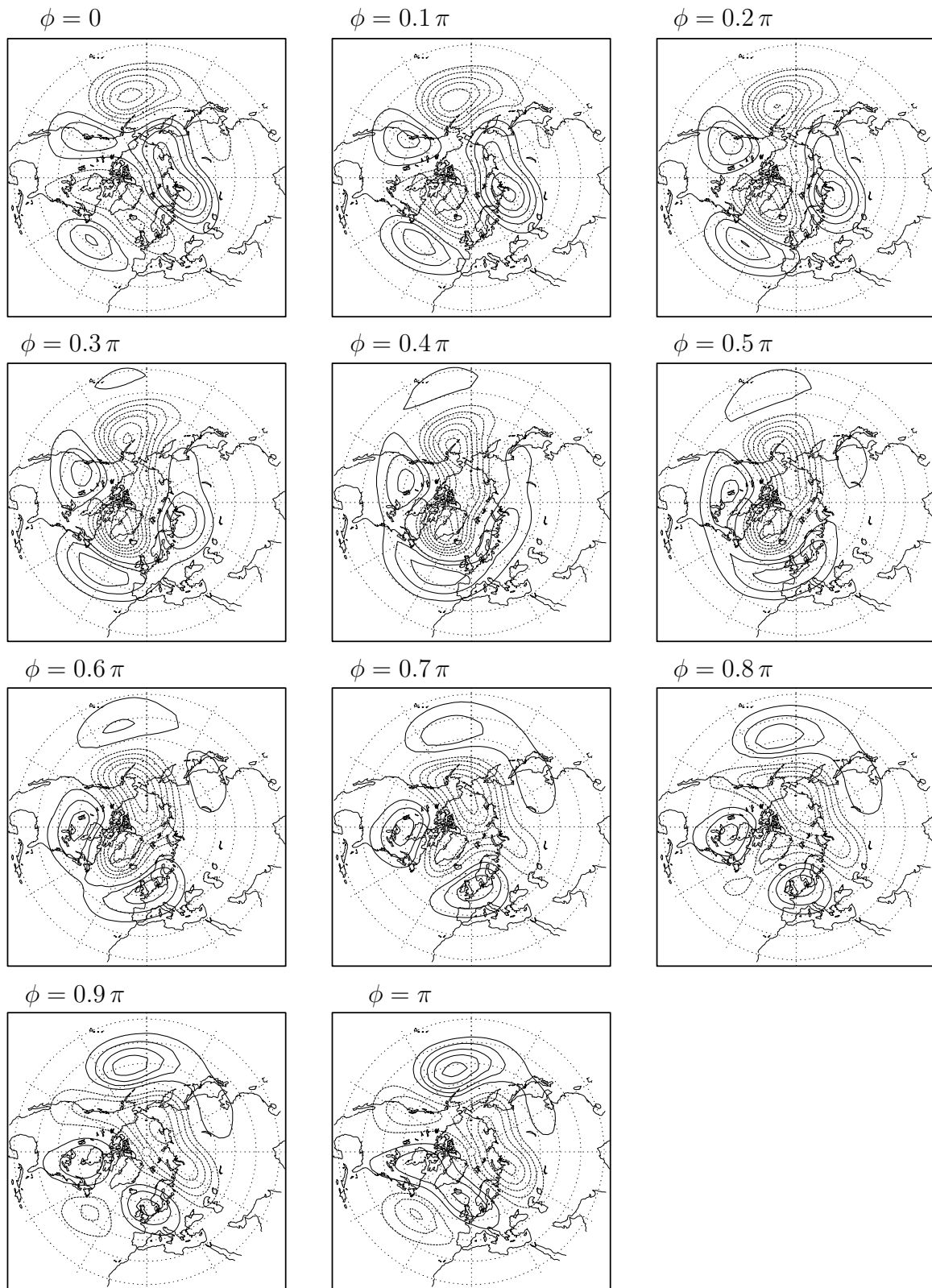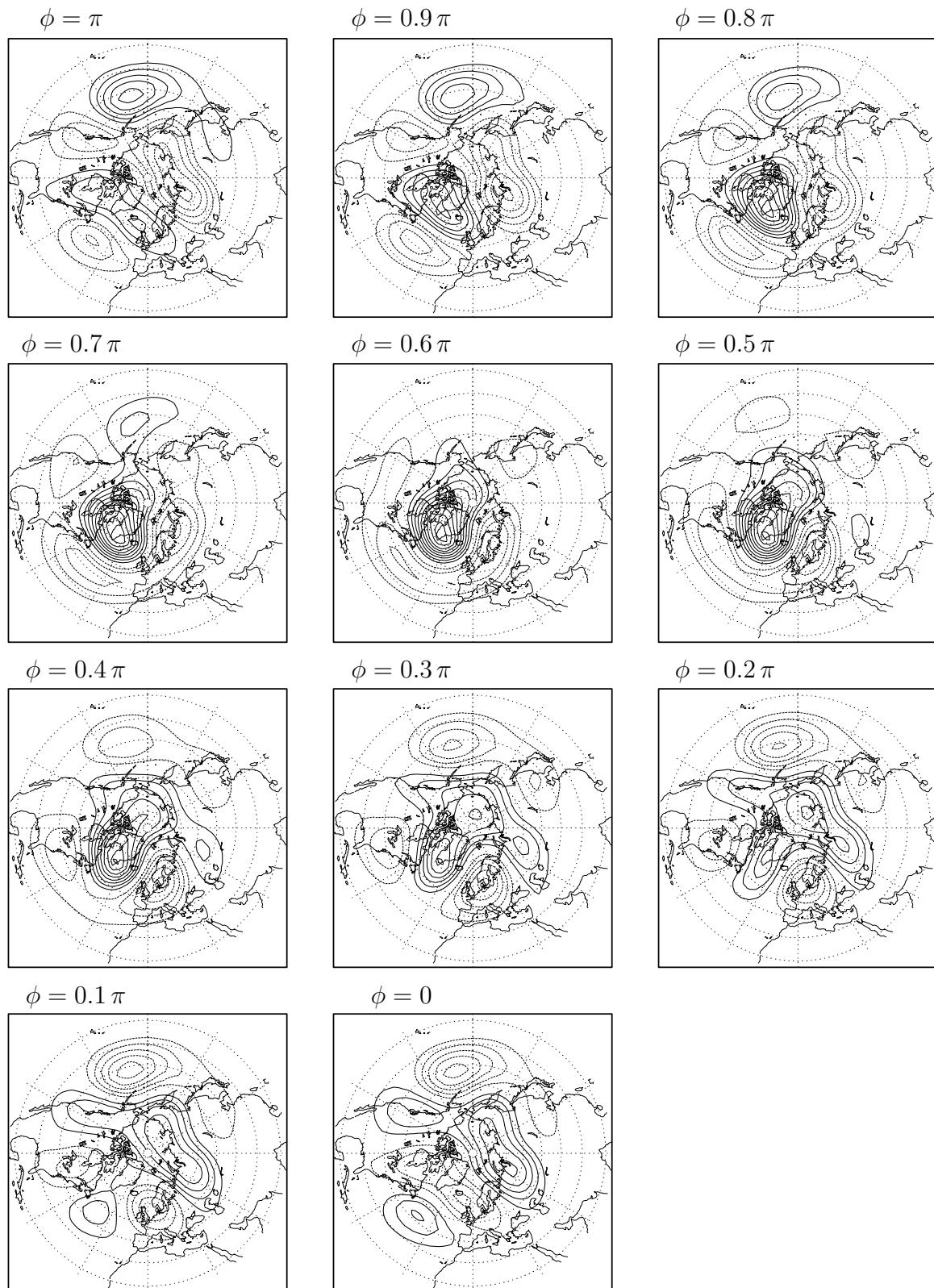
Figure 4.6: *Evolution of the atmospheric flow (500 mb geopotential height anomaly) during the second half of the preferred path through the reduced phase space $\mathcal{P}$ (see figure 4.4). Spherical coordinates are $\rho = 1$, $\theta = 1.9\,\pi$ and ϕ varying from π to 0 in steps of $0.1\,\pi$. Contour interval is 15 m.*

other values varying between 0.42 and 0.54.

The absence of maximal diagonal elements in or near other regimes than blocked NAO may have to do with the earlier mentioned problem that regimes found as PDF maxima are often spread out over several cells. Note, however, that increased persistence, or quasi-stationarity, of regimes has been detected before in observational data. Kimoto and Ghil (1993a) find increased persistence for their PNA and zonal NAO regimes when inspecting phase speeds. Their PDF of quasi-stationary states has a main maximum at the location of the blocked NAO regime, and submaxima (or ridges rather than maxima) in the other regimes. Monahan et al. (2001) find minima of the mean absolute height tendency in their A, R and G regimes.

The dynamical structure of a preferred path connecting different regimes was studied in Crommelin (2002b) in the context of a barotropic model of northern hemisphere flow. The model generated regime behaviour, involving two preferred states of mainly zonally symmetric anomalies: an arctic anomaly surrounded by reverse sign anomalies. These flow regimes were found to correspond with steady states of the model equations, in agreement with the multiple equilibria concept put forward by CDV. Furthermore, transitions between the regimes were shown to be guided by preferred paths through phase space, between the regimes. The preferred paths were interpreted as remnants of a heteroclinic cycle, a cycle of connections back and forth between steady states. In Crommelin et al. (2002) the interaction between barotropic and topographic instabilities is shown to be able to generate such a heteroclinic cycle in a simple (CDV-like) barotropic channel model. The results of the present study indicate that the dynamics associated with a heteroclinic cycle may also play a role in the regime behaviour of the real atmosphere. The assumption that only noise, due to high-frequency atmospheric motion, is responsable for regime transitions seems unable to explain the cycle of significant transition asymmetries detected here.

Following the preferred paths for regime transitions in Crommelin (2002b), a positive (negative) phase of the Arctic Oscillation (AO) is preceded by a positive (negative) phase of the North Atlantic Oscillation (NAO). Comparing with figures 4.5 and 4.6 of the present study, a similar sequence can be seen to take place along the preferred cycle in the observations. The states of mainly zonally symmetric anomalies (near $\phi \approx 0.5\pi$) are preceded by dipoles in the North-Atlantic region. These dipoles are characteristic for the NAO. Again, a dipole corresponding to a positive (negative) NAO phase is followed by a rather zonally symmetric state resembling a positive (negative) AO phase. In both model and observations, the regimes encountered along the preferred cycle consist of arctic anomalies surrounded by reverse sign anomalies (note that in the model study these regimes were interpreted as phases of the AO, whereas in observational studies they are often referred to as zonal or blocked NAO regimes. We leave this ambiguity, as we do not want to go into the heathed AO-or-NAO debate here). The episodes of the cycle with Atlantic (NAO-like) dipole anomalies have, again both in model and observations, a more transient character.

A final word on the timescale associated with the preferred path identified in the daily data. A rough estimate of the time needed to travel the complete path can be made on the basis of the transition probabilities. Since the timestep of the dataset is one day, each transition from a cell to another one takes place in one day. We have to add time

for the tendency to stick around in a cell. If the system is in cell i with "self-transition" probability M_{ii}, the expectation value for the time of residence in cell i is $\ln 0.5 / \ln M_{ii}$. In the $k = 16$ partition, the preferred path passes through 8 cells and the average value for M_{ii} is about 0.6, resulting in an expected residence time of 1.4 days for each cell. The expected total time needed for the path thus comes down to $8 \times (1 + 1.4) = 19.2$ days. For the $k = 22$ and $k = 30$ partitions the expectation times are 21.6 resp. 20.0 days.

Somewhat hidden under the noisy, diffusive behaviour of the atmospheric flow, an unstable periodic orbit can be imagined to exist which generates certain inhomogeneities and preferences in the large-scale flow dynamics. A similar conclusion is drawn by Selten and Branstator (2002) on the basis of a model study of northern hemisphere flow. The periodic orbit may well be a remnant of a heteroclinic cycle connecting large-scale flow regimes. This dynamical possiblity, also hypothesized by Kimoto and Ghil (1993b), was explored in Crommelin (2002b) and Crommelin et al. (2002). It can provide an explanation for the slowing down of the flow evolution near the regimes.

4.6   Conclusion

It has been known for some time that the distribution of observational atmospheric flow data, in a reduced phase space spanned by the leading EOFs, tends to show preferred phase space regions. Such regions, or regimes, can be hard to make visible. In this study, the projection of the data onto the unit sphere in the space of the leading (normalized) three PCs resulted in a PDF in which regions of increased probability were clearly visible. The atmospheric circulation regimes associated with these regions correspond with earlier found flow regimes. All the regimes presented in previous studies such as CW, KG1, SIG, CMP and MPF are found in the spherical PDF.

Since a probability distribution does not provide information on the dynamics of the system under study, the reduced phase space was divided into cells, all with an equal number of datapoints. The transition probability matrix M, describing (probabilities of) transitions between cells, reflects the effect of the dynamical processes governing the system behaviour, in particular the regime transitions. Contributions to M come from both diffusive, irreversible dynamics (e.g. noise-driven motion in a potential field) and conservative, reversible dynamics (generated by e.g. periodic orbits or tori in phase space). The diffusive dynamics add to the symmetric part M^S of M, whereas the antisymmetric part M^A is in principle generated by the conservative dynamics. A statistical significance test of the elements of M^A (to select elements unlikely to be due to finite sample size) left us with a set of transition asymmetries forming a closed path or cycle over the unit sphere in the reduced phase-space.

It is improbable that this preferred path can be accounted for by some form of diffusive dynamics, such as noise-induced transitions between the attraction basins of regimes. Instead, it may reflect the presence of an unstable periodic orbit in the (reduced) phase space of the large-scale atmospheric flow system. The evolution of the atmospheric flow pattern along the preferred path shows an alternation between anomaly states with considerable zonal symmetry (zonal and blocked NAO) and states with mainly meridionally oriented anomalies.

The evolution of the large-scale atmospheric flow is known to be relatively slow in the regimes, a result partially recovered in this study. The preferred path crosses the regimes of zonal and blocked NAO, and this led us to interpret the path as the manifestation of a heteroclinic cycle, a closed loop of connections in phase space between steady states associated with regimes. The role of this dynamical structure in regime behaviour was studied before in models of low order (Crommelin et al. 2002) and of moderate complexity (Crommelin, 2002b). In the latter study, regime transitions were found to be guided by remnants of heteroclinic connections. It resulted in a preferred path through phase space reminiscent of the path found in the current study, as it also connects regimes whose anomalies are dominated by zonal symmetry.

Bibliography

Achatz, U. and G. Branstator, 1999: A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *J. Atmos. Sci.*, **56,** 3140–3160.

Achatz, U. and J. D. Opsteegh, 2002. Primitive-equation-based low-order models with seasonal cycle. II. application to complexity and nonlinearity of large-scale atmosphere dynamics. To appear in *J. Atmos. Sci.*

Armbruster, D., J. Guckenheimer, and P. Holmes, 1988: Heteroclinic cycles and modulated travelling waves in systems with O(2) symmetry. *Physica D*, **29,** 257–282.

Arnol'd, V. I., 1983: *Geometrical Methods in the Theory of Ordinary Differential Equations.* Springer-Verlag.

Aubry, N., P. Holmes, J. L. Lumley, and E. Stone, 1988: The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, **192,** 115–173.

Barkmeijer, J., 1992: Local error growth in a barotropic model. *Tellus*, **44A,** 314–323.

Baur, F., 1947: *Musterbeispiele Europäischer Großwetterlagen.* Dieterich, 35 pp.

Benzi, R., A. R. Hansen, and A. Sutera, 1984: On stochastic perturbations of simple blocking models. *Quart. J. Roy. Meteor. Soc.*, **110,** 393–409.

Benzi, R., P. Malguzzi, A. Speranza, and A. Sutera, 1986: The statistical properties of general atmospheric circulation: Observational evidence and a minimal theory of bimodality. *Quart. J. Roy. Meteor. Soc.*, **112,** 661–674.

Blender, R. and P. Névir, 1991: Nonlinear wave and modon solutions on the β-plane generated by Lie-symmetries. *Beitr. Phys. Atmosph.*, **64,** 255–258.

Branstator, G. and J.D. Opsteegh, 1989: Free solutions of the barotropic vorticity equation. *J. Atmos. Sci.*, **46,** 1799–1814.

Broer, H. W., C. Simo, and R. Vitolo, 2002: Bifurcations and strange attractors in the Lorenz-84 climate model with seasonal forcing. *Nonlinearity*, **15,** 1205–1267.

Charney, J. G. and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36,** 1205–1216.

Charney, J. G., J. Shukla, and K. C. Mo, 1981: Comparison of a barotropic blocking theory with observation. *J. Atmos. Sci.*, **38,** 762–779.

Cheng, X. and J. M. Wallace, 1993: Cluster analysis of the northern hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50,** 1205–1216.

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

Crommelin, D. T., 2002a: Homoclinic dynamics: a scenario for atmospheric ultra-low-frequency variability. *J. Atmos. Sci.*, **59,** 1533–1549.

Crommelin, D. T., 2002b. Observed non-diffusive dynamics in large-scale atmospheric flow. Submitted to *J. Atmos. Sci.*

Crommelin, D. T., 2002c. Regime transitions and heteroclinic connections in a barotropic atmosphere. To appear in *J. Atmos. Sci.*

Crommelin, D. T., J. D. Opsteegh, and F. Verhulst, 2002. A mechanism for atmospheric regime behaviour. Submitted to *J. Atmos. Sci.*

Cushman-Roisin, B., 1994: *Introduction to Geophysical Fluid Dynamics.* Prentice-Hall, 320 pp.

De Swart, H. E., 1988a: Low-order spectral models of the atmospheric circulation: a survey. *Acta Appl. Math.*, **11,** 49–96.

De Swart, H. E., 1988b: *Vacillation and Predictability Properties of Low-Order Atmospheric Spectral Models.* PhD thesis, Rijksuniversiteit Utrecht.

De Swart, H. E., 1989: Analysis of a six-component atmospheric spectral model: Chaos, predictability and vacillation. *Physica D*, **36,** 222–234.

De Swart, H. E. and J. Grasman, 1987: Effect of stochastic perturbations on a low-order spectral model of the atmospheric circulation. *Tellus*, **39A,** 10–24.

Dethloff, K., A. Weisheimer, A. Rinke, D. Handorf, M.V. Kurgansky, W. Jansen, P. Maaß, and P. Hupfer, 1998: Climate variability in a nonlinear atmosphere-like dynamical system. *J. Geophys. Res.*, **103,** 25957–25966.

Doedel, E., 1986. AUTO: Software for continuation and bifurcation problems in ordinary differential equations. California Institute of Technology.

Doedel, E. J. and M. J. Friedman, 1989: Numerical computation of heteroclinic orbits. *J. Comput. Appl. Math.*, **26,** 155–170.

Doedel, E. J. and X. J. Wang, 1995: AUTO94: Software for continuation and bifurcation problems in ordinary differential equations. Technical Report CRPC-95-2, Center for Research on Parallel Computing, California Institute of Technology.

Egger, J., 1981: Stochastically driven large-scale circulations with multiple equilibria. *J. Atmos. Sci.*, **38,** 2606–2618.

Feller, W., 1968: *An Introduction to Probability Theory and Its Applications, Volume I, 3rd edition.* John Wiley & Sons, 509 pp.

Fowler, A.C. and C.T. Sparrow, 1991: Bifocal homoclinic orbits in four dimensions. *Nonlinearity*, **4,** 1159–1182.

Ghil, M. and S. Childress, 1987: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics.* Springer-Verlag.

Glendinning, P., 1997: Differential equations with bifocal homoclinic orbits. *Int. J. Bifurcation and Chaos*, **7,** 27–37.

Guckenheimer, J. and P. Holmes, 1983: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag, 459 pp.

Haines, K., 1994: Low-frequency variability in atmospheric middle latitudes. *Surveys in Geophysics*, **15,** 1–61.

Handorf, D., V.K. Petoukhov, K. Dethloff, A.V. Eliseev, A. Weisheimer, and I.I. Mokhov, 1999: Decadal climate variability in a coupled atmosphere-ocean climate model of moderate complexity. *J. Geophys. Res.*, **104,** 27253–27275.

Holmes, P., J. L. Lumley, and G. Berkooz, 1996: *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, 420 pp.

Holmes, P.J., J.L. Lumley, G. Berkooz, J.C. Mattingly, and R.W. Wittenberg, 1997: Low-dimensional models of coherent structures in turbulence. *Phys. Rep.*, **287,** 337–384.

Holton, J. R., 1992: *An Introduction to Dynamic Meteorology, 3rd edition.* Academic Press, 511 pp.

Itoh, H. and M. Kimoto, 1996: Multiple attractors and chaotic itinerancy in a quasi-geostrophic model with realistic topography: Implications for weather regimes and low-frequency variability. *J. Atmos. Sci.*, **53,** 2217–2231.

Itoh, H. and M. Kimoto, 1997: Chaotic itinerancy with preferred transition routes appearing in an atmospheric model. *Phys. D*, **109,** 274–292.

Itoh, H. and M. Kimoto, 1999: Weather regimes, low-frequency oscillations, and principal patterns of variability: A perspective of extratropical low-frequency variability. *J. Atmos. Sci.*, **56,** 2684–2705.

James, I.N. and P.M. James, 1989: Ultra-low-frequency variability in a simple atmospheric circulation model. *Nature*, **342,** 53–55.

James, I.N. and P.M. James, 1992: Spatial structure of ultra-low-frequency variability of the flow in a simple atmospheric circulation model. *Quart. J. R. Met. Soc.*, **118,** 1211–1233.

James, P.M., K. Fraedrich, and I.N. James, 1994: Wave-zonal-flow interaction and ultra-low-frequency variability in a simplified global circulation model. *Quart. J. R. Met. Soc.*, **120,** 1045–1067.

Källén, E., 1981: The nonlinear effects of orographic and momentum forcing in a low-order, barotropic model. *J. Atmos. Sci.*, **38,** 2150–2163.

Källén, E., 1982: Bifurcation properties of quasi-geostrophic, barotropic models and their relation to atmospheric blocking. *Tellus*, **34,** 255–265.

Kimoto, M. and M. Ghil, 1993a: Multiple flow regimes in the northern hemisphere winter. part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, **50,** 2625–2643.

Kimoto, M. and M. Ghil, 1993b: Multiple flow regimes in the northern hemisphere winter. part II: Sectorial regimes and preferred transitions. *J. Atmos. Sci.*, **50,** 2645–2673.

Knobloch, E. and J. Moehlis, 2000. Burst mechanisms in hydrodynamics. In *Nonlinear Instability, Chaos and Turbulence, Volume II*, ed. L. Debnath and D. Riahi, Computational Mechanics Publications, Southampton, 237-287.

Krupa, M., 1997: Robust heteroclinic cycles. *J. Nonlinear Sci.*, **7,** 129–176.

Kuo, H. L., 1949: Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteorol.*, **6,** 105–122.

Kurgansky, M.V., K. Dethloff, I.A. Pisnichenko, H. Gernandt, F.-M. Chmielewski, and W. Jansen, 1996: Long-term climate variability in a simple, nonlinear atmospheric model. *J. Geophys. Res.*, **101,** 4299–4314.

Kuznetsov, Yu. A., 1995: *Elements of Applied Bifurcation Theory*. Springer-Verlag, 515 pp.

Kuznetsov, Yu. A., 1999: Numerical normalization techniques for all codim 2 bifurcations of equilibria in ODE's. *SIAM J. Numer. Anal.*, **36,** 1104–1124.

Kuznetsov, Yu. A. and V. V. Levitin, 1997. CONTENT: A multiplatform environment for analyzing dynamical systems. Dynamical Systems Laboratory, Centrum voor Wiskunde en Informatica, Amsterdam, ftp.cwi.nl/pub/CONTENT.

Laing, C. and P. Glendinning, 1997: Bifocal homoclinic bifurcations. *Physica D*, **102,** 1–14.

Lau, K.-M. and L. Peng, 1987: Origin of low-frequency (intraseasonal) oscillations in the tropical atmosphere. part i: Basic theory. *J. Atmos. Sci.*, **44,** 950–972.

LeBlanc, V. G., 2000: On some secondary bifurcations near resonant Hopf-Hopf interactions. *Dyn. Cont. Discr. Imp. Syst.*, **7,** 405–427.

Legras, B. and M. Ghil, 1985: Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.*, **42,** 433–471.

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus*, **12,** 364–373.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20,** 130–141.

Lorenz, E. N., 1984: Irregularity: A fundamental property of the atmosphere. *Tellus*, **36A,** 98–110.

Mo, K. C. and M. Ghil, 1987: Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.*, **44,** 877–901.

Mo, K. C. and M. Ghil, 1988: Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.*, **93,** 10927–10952.

Molteni, F., S. Tibaldi, and T. N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Quart. J. Roy. Meteor. Soc.*, **116,** 31–67.

Monahan, A. H., L. Pandolfo, and J. C. Fyfe, 2001: The preferred structure of variability of the northern hemisphere atmospheric circulation. *Geophys. Res. Lett.*, **28,** 1019–1022.

Namias, J., 1950: The index cycle and its role in the general circulation. *J. Meteorol.*, **7,** 130–139.

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

Pasmanter, R. A. and A. Timmermann, 2002. Cyclic Markov chains with an application to ENSO predictability. To appear in *Nonlin. Proc. Geophys.*

Pielke, R.A. and X. Zeng, 1994: Long-term variability of climate. *J. Atmos. Sci.*, **51,** 155–159.

Plaut, G. and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *J. Atmos. Sci.*, **51,** 210–236.

Preisendorfer, R. W., 1988. *Principal Component Analysis in Meteorology and Oceanography.* ed. C.D. Mobley. Elsevier, Amsterdam, 401 pp.

Proctor, M. R. E. and C. A. Jones, 1988: The interaction of two spatially resonant patterns in thermal convection. Part 1. Exact 1:2 resonance. *J. Fluid Mech.*, **188,** 301–335.

Reinhold, B. B. and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.*, **110,** 1105–1145.

Rex, D.F., 1950a: Blocking action in the middle troposphere and its effect on regional climate. I. an aerological study of blocking action. *Tellus*, **2,** 196–211.

Rex, D.F., 1950b: Blocking action in the middle troposphere and its effect on regional climate. II. the climatology of blocking action. *Tellus*, **2,** 275–301.

Roads, J. O., 1987: Predictability in the extended range. *J. Atmos. Sci.*, **44,** 3495–3527.

Salby, M.L. and H.H. Hendon, 1994: Intraseasonal behaviour of clouds, temperature, and motions in the tropics. *J. Atmos. Sci.*, **51,** 2207–2224.

Selten, F. M., 1995: An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.*, **52,** 915–936.

Selten, F. M. and G. Branstator, 2002. An unstable periodic orbit and preferred regime transition routes in a baroclinic model. In preparation.

Shil'nikov, A., G. Nicolis, and C. Nicolis, 1995: Bifurcation and predictability analysis of a low-order atmospheric circulation model. *Int. J. Bifurcation and Chaos*, **5,** 1701–1711.

Shil'nikov, L. P., 1965: A case of the existence of a denumerable set of periodic motions. *Sov. Math. Dokl.*, **6,** 163–166.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 175 pp.

Smale, S., 1963. Diffeomorphisms with many periodic points. In *Differential and Combinatorial Topology*, ed. S. S. Cairns, Princeton University Press, Princeton, pp. 63-80.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in northern hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56,** 3704–3723.

Spekat, A., B. Heller-Schulze, and M. Lutz, 1983: Über Grosswetter und Markov Ketten. *Meteor. Rundsch.*, **36,** 243–248.

Sura, P., 2002: Noise-induced transitions in a barotropic β-plane channel. *J. Atmos. Sci.*, **59,** 97–110.

Sutera, A., 1986: Probability density distribution of large-scale atmospheric flow. *Adv. Geophys.*, **29,** 227–249.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part 1: Theory. *Q. J. R. Meteorol. Soc.*, **113,** 1311–1328.

Thompson, D. W. J. and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25,** 1297–1300.

Tian, Y., E. R. Weeks, K. Ide, J. S. Urbach, C. N. Baroud, M. Ghil, and H. L. Swinney, 2001: Experimental studies of an eastward jet over topography. *J. Fluid Mech.*, **438,** 129–157.

Tung, K. K. and A. J. Rosenthal, 1985: Theories of multiple equilibria - a critical reexamination. part I: Barotropic models. *J. Atmos. Sci.*, **42,** 2804–2819.

Tuwankotta, J. M., 2002: *Higher Order Resonances in Dynamical Systems*. PhD thesis, Universiteit Utrecht.

Van Veen, L., 2002a. Baroclinic flow and the Lorenz-84 model. To appear in *Inter. J. Bifur. Chaos*.

Van Veen, L., 2002b: *Time Scale Interaction in Low-Order Climate Models*. PhD thesis, Universiteit Utrecht.

Vautard, R., K. C. Mo, and M. Ghil, 1990: Statistical significance test for transition matrices of atmospheric Markov chains. *J. Atmos. Sci.*, **47,** 1926–1931.

Vickroy, J. G. and J. A. Dutton, 1979: Bifurcation and catastrophe in a simple, forced, dissipative quasi-geostrophic flow. *J. Atmos. Sci.*, **36,** 42–52.

Von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, 1988: Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere. *J. Geophys. Res.*, **93,** 11022–11036.

Wallace, J. M. and M. L. Blackmon, 1983. Observations of low-frequency atmospheric variability. In *Large-scale Dynamical Processes in the Atmosphere*, ed. B. J. Hoskins and R. P. Pearce, Academic Press, London, pp. 55-94.

Wallace, J. M. and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Norhtern Hemisphere winter. *Mon. Wea. Rev.*, **109,** 784–812.

Weeks, E. R., Y. Tian, J. S. Urbach, K. Ide, H. L. Swinney, and M. Ghil, 1997: Transitions between blocked and zonal flows in a rotating annulus with topography. *Science*, **278,** 1598–1601.

Wiggins, S., 1988: *Global Bifurcations and Chaos*. Springer-Verlag, 494 pp.

Yoden, S., 1985a: Bifurcation properties of a quasi-geostrophic, barotropic, low-order model with topography. *J. Meteor. Soc. Japan*, **63,** 535–546.

Yoden, S., 1985b: Multiple steady states of quasi-geostrophic barotropic flow over sinusoidal topography. *J. Meteor. Soc. Japan*, **63,** 1031–1045.

Samenvatting

Het gedrag van de grootschalige luchtstroming in de atmosfeer wordt gedomineerd door het bestaan van een klein aantal voorkeurspatronen voor de atmosferische circulatie. Zulke voorkeurspatronen heten *regimes*. Veranderingen in de atmosferische stroming op een ruimtelijke schaal van duizenden kilometers, op een tijdschaal van een week tot een paar maanden, bestaan voor een belangrijk deel uit zogeheten regime-gedrag: de atmosfeer is tijdelijk "gevangen" in een regime, maakt een transitie naar een ander regime, verblijft daar ook weer enige tijd, maakt weer een transitie, enzovoorts. De grootschalige stroming bepaalt hoe weersystemen (bijv. depressies) zich bewegen, en is daarmee van belang voor het weertype in bijvoorbeeld Nederland. Bevindt de atmosfeer zich gedurende de winter bijvoorbeeld in een regime met *zonale* (d.w.z. oostwaarts gerichte) stroming boven de Atlantische Oceaan, dan zal het weer in Nederland overwegend zacht en regenachtig zijn, omdat vooral relatief warme en vochtige lucht van boven de oceaan aangevoerd wordt. Gedurende een *geblokkeerd* regime daarentegen, waarin de west-oost stroming (de straalstroom) boven de oceaan geblokkeerd is door een groot hogedrukgebied en daarom deels "omgeleid" wordt over Groenland en Scandinavië, zal het in Nederland eerder koud en droog zijn door de aanvoer van polaire lucht.

De laatste jaren zijn aanwijzingen gevonden dat mondiale klimaatveranderingen samenhangen met veranderingen in het regime-gedrag. De regimes zelf veranderen niet zozeer, maar wel de frequentie waarmee het atmosfeersysteem ze bezoekt. In de jaren vijftig en zestig van de 20^e eeuw kwam het hiervoor beschreven geblokkeerde regime bijvoorbeeld vaker voor dan in de decennia daarna. In dit licht valt te begrijpen waarom men verwacht dat de huidige en toekomstige klimaatveranderingen ten gevolge van het broeikaseffect zullen resulteren in zachtere en nattere winters in noord-west Europa.

Onderzoek van langjarige reeksen van meetgegevens van de atmosferische circulatie is redelijk succesvol geweest in het identificeren van regimes. Echter, een eenduidige verklaring van de mechanismen en processen die regime-gedrag veroorzaken ontbreekt nog. Een eerste belangrijke stap werd gezet in een artikel uit 1979 van Charney en DeVore. Zij suggereerden dat regimes geassocieerd zijn met *evenwichtstoestanden* van de grootschalige atmosferische stroming. De wiskundige vergelijkingen die het gedrag van de grootschalige stroming beschrijven hebben wellicht evenwichtsoplossingen die overeenkomen met regimes. Ze lieten zien dat dit inderdaad het geval is in een sterk vereenvoudigd dynamisch model voor de atmosfeer. Het feit dat dit model meerdere evenwichtsoplossingen heeft voor een en dezelfde instelling van de fysische parameters (bijv. wrijving), heeft te maken met de aanwezigheid van gebergten ("orografie", of "topografie") in het model en met het *niet-lineaire* karakter van de modelvergelijkingen.

De wiskundige vergelijkingen voor de atmosferische stroming zijn afgeleid van de

Navier-Stokes vergelijkingen: de algemene natuurwet die het stromingsgedrag van gassen en vloeistoffen beschrijft. De Navier-Stokes vergelijkingen, en daarmee alle eruit afgeleide vergelijkingen voor speciale stromingssystemen zoals de atmosfeer, zijn niet-lineair. Dat wil zeggen dat er allerlei (positieve en negatieve) terugkoppelingseffecten kunnen optreden. Het resultaat van die terugkoppelingsmechanismen is meestal moeilijk te doorgronden. Een van de mogelijkheden is dat het systeem zich *chaotisch*, d.w.z. intrinsiek onvoorspelbaar, gedraagt. De atmosfeer is, met aan zekerheid grenzende waarschijnlijkheid, chaotisch (een keihard wiskundig bewijs is erg moeilijk te geven), met als gevolg o.a. de beperkte geldigheid van weersvoorspellingen. Het deelgebied van de wiskunde dat zich bezig houdt met niet-lineaire systemen, is dat der *dynamische systemen*, ook wel niet-lineaire dynamica (en in populair-wetenschappelijke artikelen soms "chaostheorie") genoemd. In dit proefschrift wordt het regime-gedrag van de atmosfeer geanalyseerd aan de hand van technieken en concepten uit de theorie van dynamische systemen.

De associatie van regimes met stabiele evenwichtsoplossingen, zoals voorgesteld door Charney en DeVore, werpt nog geen licht op de overgangen tussen de regimes. Immers, als het atmosfeersysteem zich in stabiel evenwicht zou bevinden, vergelijkbaar met een slinger die recht naar beneden hangt, kan het dit evenwicht niet meer verlaten. Meestal is over regime-transities nagedacht als veroorzaakt door *ruis*: kleinschalige processen in de atmosfeer werken effectief als ruis op de grootschalige stroming, waardoor het grootschalige evenwicht verloren kan gaan. Echter, het is ook mogelijk dat de grootschalige stroming eigenhandig, zonder de invloed van ruis, regime-overgangen kan maken. De transities worden in dat geval voortgebracht door de interne niet-lineaire dynamica van de grootschalige atmosferische stroming. Deze tweede mogelijkheid wordt in dit proefschrift uitgewerkt.

Het concept uit de theorie van dynamische systemen dat de transities tussen voorkeurstoestanden in essentie zou kunnen beschrijven, is dat van de zogenaamde *heterocliene connecties*. Dat zijn verbindingen tussen evenwichtstoestanden in de toestandsruimte (een "ruimte" waarin ieder punt overeenkomt met een aparte toestand van de atmosferische stroming). Als regime-overgangen inderdaad te maken hebben met dergelijke heterocliene connecties, zullen de overgangen niet geheel willekeurig verlopen maar daarentegen bepaalde voorkeursbanen volgen door de toestandsruimte.

In hoofdstuk een wordt de algemene dynamica van een eenvoudig atmosfeermodel geanalyseerd. De ideeën die hieruit voortkwamen gaven de inspiratie tot de specifieke benadering en analyse van het regime-gedrag, die in de hoofdstukken twee, drie en vier centraal staan. Statistische analyse van transities in een model (hoofdstuk twee) en in de echte atmosfeer (hoofdstuk vier) wijst uit dat de transities inderdaad voorkeursbanen vertonen. In de echte atmosfeer zijn die voorkeuren moeilijk zichtbaar te maken omdat ze vertroebeld worden door de aanwezige ruis. In het in hoofdstuk twee geanalyseerde atmosfeermodel zijn de voorkeuren duidelijk te zien. Een benadering (m.b.v. de computer) van de heterocliene connecties in dit model laat een goede overeenkomst zien tussen de gevonden connecties en de statistisch bepaalde voorkeuren van de transities. In hoofdstuk drie wordt een sterk vereenvoudigd model geanalyseerd om meer duidelijkheid te krijgen omtrent het wiskundige scenario dat de heterocliene connecties mogelijk maakt. De aanwezigheid van zulke connecties is namelijk niet vanzelfsprekend. Echter, twee elementen die altijd in atmosfeermodellen voorkomen (de aanwezigheid van bergen, en een "instabiele" straalstroom) blijken voldoende te zijn om de connecties mogelijk te maken.

Dankwoord

Na een paar jaar het slappe koord tussen wiskunde en meteorologie, tussen KNMI en Mathematisch Instituut te hebben bewandeld, is het moment daar om diegenen te bedanken die mij hebben geholpen niet van het koord af te donderen:

Theo en Ferdinand, voor hun begeleiding en vertrouwen, en voor de ruimte die ze me gaven om mijn ondoordachte ideeën, verleidelijke dwaalwegen en vruchtbare ingevingen na te jagen.

Lennaert, medewandelaar, van wie ik menig balanceertruck heb afgekeken.

Gerard, Wim, Rubén, Frank, Axel, Ulrich, Rein, Erik, Nanne, Pedro, Hanneke, Casper, Jan, Xueli, Greet, Brigitta, en alle andere KNMI-collegae, die mij met een uniek mengsel van koffie en klimaatonderzoek altijd weer wisten te lokken uit het woud der wiskunde.

Ulrich and Grant are kindly thanked for the use of their model, and for their hospitality when I visited their institutes.

Bob, Theo, Menno, Mischja, Luis, Barbara, Philip, Arno, Abadi, Siti, Taoufik, Martijn, Yuri, Odo, Thijs, Geertje, die een fysicus zich thuis deden voelen op een Mathematisch Instituut (daarbij geholpen door menig Bastaard-sessie).

Mijn ouders, voor hun steun, hun belangstelling en hun genen.

Ernestine, voor de zon in Oost, het leven op Borneo, de muziek in mijn hoofd. En voor haar onnavolgbare kijk op het Shil'nikov fenomeen.

Curriculum Vitae

Daan Crommelin werd geboren op 10 augustus 1972 te Delft, doch groeide op in Zoetermeer. In 1990 haalde hij zijn VWO-diploma aan het Alfrink College te Zoetermeer en studeerde vervolgens een jaar aan het Institut Savoisien d'Etudes Françaises in Chambéry, Frankrijk. In 1991 begon hij aan de studie natuurkunde aan de Universiteit van Amsterdam, vanaf 1992 aangevuld met de studie wijsbegeerte aan dezelfde universiteit. Hij was student-lid van onderwijscommissie en faculteitsraad van de faculteit Natuur- en Sterrenkunde. Afgestudeerd in 1997 in de richtingen theoretische natuurkunde alsmede systematische en historische wijsbegeerte, startte hij in januari 1998 met een promotieonderzoek, uitgevoerd aan het KNMI en het Mathematisch Instituut van de Universiteit Utrecht. De wegen der wetenschap leidden hem in de afgelopen jaren naar conferenties, zomerscholen en werkbezoeken in Reading, Nice, Enschede, Snowbird, Boulder, Boekelo, l'Aquila, Berg en Dal, Kühlungsborn en Kerkrade. Het onderzoek resulteerde in het proefschrift dat de lezer nu in handen houdt.